BGRS/SB-2024

БИОИНФОРМАТИКА РЕГУЛЯЦИИ И СТРУКТУРЫ ГЕНОМОВ / СИСТЕМНАЯ БИОЛОГИЯ

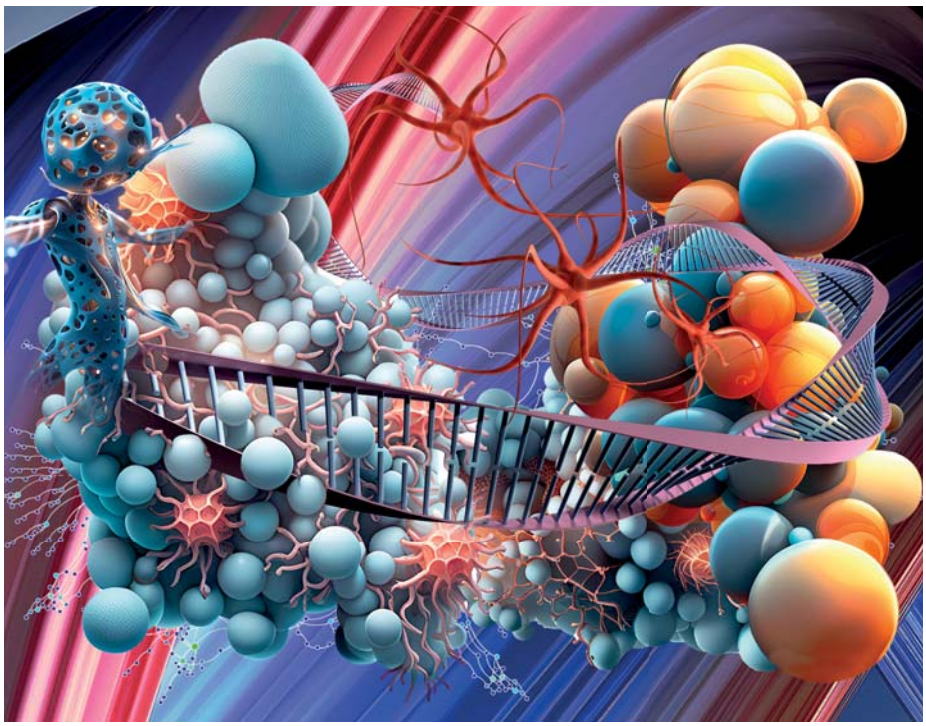
14-я Международная мультиконференция

5–10 августа 2024, Новосибирск, Россия

BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE / SYSTEMS BIOLOGY

14th International Multiconference

August 5–10, 2024, Novosibirsk, Russia



https://bgrssb.icgbio.ru/2024/

Организаторы/Organizers

Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук» (ИЦиГ СО РАН)
Institute of Cytology and Genetics, SB RAS (ICG SB RAS)

Новосибирский государственный университет (НГУ)
Novosibirsk State University (NSU)

Научный совет по биоинформатике СО РАН
Scientific Council on Bioinformatics SB RAS

Спонсоры/Sponsors

Конференция поддерживается Курчатовским геномным центром
Института цитологии и генетики СО РАН (номер проекта 075-15-2019-1662)
The conference is supported by the Kurchatov Genomic Center
of the Institute of Cytology and Genetics, SB RAS (Novosibirsk)
(grant number: 075-15-2019-1662)

ЗОЛОТЫЕ СПОНСОРЫ/GOLD SPONSORS

ООО «БИОЛАБМИКС»/ BIOLABMIX
https://biolabmix.ru/

ООО «ДИАЭМ»/ DIA-M
www.dia-m.ru

ООО «СкайДжин»/ SkyGen
www.skygen.com

СЕРЕБРЯНЫЕ СПОНСОРЫ/SILVER SPONSORS

ООО «НПФ СИНТОЛ»/ SINTOL
https://www.syntol.ru/

ООО «МПБА ДИАГНОСТИКА»/ MP Biomedicals
http://www.mpbio.ru/

ООО «КОМПАНИЯ ХЕЛИКОН»/ HELICON
https://helicon.ru/

БРОНЗОВЫЕ СПОНСОРЫ/BRONZE SPONSORS

ООО «Альгавитапро»/ NextGenSeq.tech
https://nextgenseq.tech/

ООО «БелБиоЛаб»/ BelBioLab
https://belbiolab.ru/

ООО «ТД «ХИММЕД»/ CHIMMED
https://chimmed.ru/

ООО «ЛАКОПА»/ LLC «Lacopa»
https://lacopa.group/

ООО «СЕСАНА»/SESANA
https://sesana.ru/

БАЗОВЫЙ СПОНСОР/ BASIC SPONSOR

ООО «БХБ»/ Biohim Store
https://www.biohim.store/

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук»
Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет»

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk State University

БИОИНФОРМАТИКА РЕГУЛЯЦИИ И СТРУКТУРЫ ГЕНОМОВ / СИСТЕМНАЯ БИОЛОГИЯ

Четырнадцатая международная мультиконференция

Тезисы докладов

5–10 августа 2024 г., Новосибирск, Россия

BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE / SYSTEMS BIOLOGY (BGRS/SB-2024)

The Fourteenth International Multiconference

Abstracts

August 5–10, 2024, Novosibirsk, Russia

Новосибирск
ИЦиГ СО РАН
2024

Б63

Биоинформатика регуляции и структуры геномов / системная биология: Четырнадцатая международная мультиконференция : Тез. докл. (Россия, Новосибирск, 5–10 авг. 2024 г.) / Ин-т цитологии и генетики СО РАН; Новосиб. нац. исслед. гос. ун-т. – Новосибирск: ИЦиГ СО РАН, 2024. – XXXVI с., 2372 с. – ISBN 978-5-91291-067-8

Bioinformatics of Genome Regulation and Structure / Systems Biology (BGRS/SB-2024) : The Fourteenth International Multiconference (August 5–10, 2024, Novosibirsk, Russia) : Abstracts / Inst. of Cytol. and Genetics, Sib. Branch of the Russ. Acad. of Sci.; Novosib. State Univ. – Novosibirsk: ICG SB RAS, 2024. – XXXVI p., 2372 p. – ISBN 978-5-91291-067-8

DOI 10.18699/bgrs2024-abstracts

Организационный комитет

Институт цитологии и генетики СО РАН:

Зубова Светлана Васильевна, рук. сектора – председатель оргкомитета
Антропова Евгения Александровна, м.н.с.
Ахметова Веста Дарвиновна, зам. нач. отдела
Батухтин Георгий Валерьевич, редактор
Иванов Роман Артемович, м.н.с., программист
Игнатьева Ольга Валерьевна, юрисконсульт
Карамышева Татьяна Витальевна, с.н.с., к.б.н.
Коваль Василий Сергеевич, вед. специалист, к.б.н.
Линкевич Павел Евгеньевич, вед. инж.-программист
Морковина Алина Владимировна, программист
Николаев Александр Олегович, вед. экономист
Савватеева Ольга Николаевна, вед. специалист
Стукова Анастасия Владимировна, вед. экономист
Токпанов Ерлан Аскарович, рук. службы инф. тех.
Харкевич Андрей Владимирович, вед. специалист
Чалкова Татьяна Федоровна, нач. отдела
Черкавский Андрей Дмитриевич, программист

Organizing committee

Institute of Cytology and Genetics, SB RAS:

Zubova Svetlana (Chairperson)
Akhmetova Vesta
Antropova Evgenia
Batukhtin Georgy
Chalkova Tatiana
Cherkavsky Andrey
Ignatieva Olga
Ivanov Roman
Karamysheva Tatiana
Kharkevich Andrey
Koval Vasily
Linkevich Pavel
Morkovina Alina
Nikolaev Alexander
Savvateeva Olga
Stukova Anastasia
Tokpanov Erlan

Контакты

Официальный адрес оргкомитета конференции: bgrs2024@bionet.nsc.ru
Адрес: 630090, Новосибирск, проспект Академика Лаврентьева, 10, ИЦиГ СО РАН
Председатель оргкомитета: С.В. Зубова, svetazubova@gmail.com, +7 (383) 363 4977

Contacts

All questions, please contact the official address
of the Organizing committee of the conference: bgrs2024@bionet.nsc.ru
Organizing committee Chairman: Zubova Svetlana
Address: Lavrentieva Avenue, 10, Novosibirsk, 630090
Phone: +7 (383) 363 4977, email: svetazubova@gmail.com

ISBN 978-5-91291-067-8

Президиум мультиконференции BGRS/SB-2024 включает следующих сопредседателей:

Колчанов Николай Александрович, академик РАН, научный руководитель ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Кочетов Алексей Владимирович, академик РАН, директор ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Федорук Михаил Петрович, академик РАН, ректор НГУ, Новосибирск, Россия

Соколов Игорь Анатольевич, академик РАН, заместитель академика-секретаря ОНИТ РАН, руководитель Секции информационных технологий и автоматизации, декан Факультета вычислительной математики и кибернетики МГУ, Москва, Россия

Хофештадт Ральф, профессор университета Билефельда, Германия (вместе с Н.А. Колчановым – организатор всех конференций BGRS/SB с самого начала, с 1998 г.)

Чень Мин, профессор, Университет Чжэцзян, Ханчжоу, Китай

Секретариат мультиконференции (ИЦиГ СО РАН, Новосибирск, Россия):

Антропова Евгения Александровна – ученый секретарь конференции

Бухарина Татьяна Анатольевна

Павлова Наталия Павловна

Программный комитет

СИМПОЗИУМ 1. Геномика, транскриптомика и биоинформатика

Секция 1.1. Структурно-функциональная организация геномов

Ершов Никита Игоревич, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Афонников Дмитрий Аркадьевич, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Секция 1.2. Регуляторная геномика

Кулаковский Иван Владимирович, Институт белка РАН, Пущино; Институт общей генетики им. Н.И. Вавилова РАН, Москва; Казанский федеральный университет

Макеев Всеволод Юрьевич, Институт общей генетики им. Н.И. Вавилова РАН, Московский физико-технический институт, Москва, Россия

Левицкий Виктор Георгиевич, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Меркулова Татьяна Ивановна, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Секция 1.3. Фундаментальная и прикладная 3D геномика

Графодатский Александр Сергеевич, Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия

Вениамин Семенович Фишман, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Баттулин Нариман Рашитович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Секция 1.4. Полногеномная транскриптомика (дифференциальная экспрессия генов)

Ощепков Дмитрий Юрьевич, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Пономаренко Михаил Павлович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

СИМПОЗИУМ 2. Системная компьютерная биология

Секция 2.1. Реконструкция, компьютерный анализ и моделирование генных сетей и метаболических путей

Барберис Маттео, Университет Суррея, Гилфорд, Великобритания

Лашин Сергей Александрович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Говорун Вадим Маркович, НИИ системной биологии и медицины Роспотребнадзора, Москва, Россия

Акбердин Илья Ринатович, Биософт; ФИЦ Институт цитологии и генетики СО РАН, Новосибирский государственный университет, Новосибирск, Россия; Университет «Сириус», Сочи, Россия

Иванисенко Владимир Александрович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Минг Чен, Чжэцзянский университет, Чжэцзянь, Китай

Хофештадт Ральф, Университет Билефельда, Германия

Секция 2.2. Компьютерный анализ и моделирование популяционных, экологических и генетических систем и процессов

Фрисман Ефим Яковлевич, Институт комплексного анализа региональных проблем ДВО РАН, Биробиджан, Россия

Жданова Оксана Леонидовна, Институт автоматики и процессов управления ДВО РАН, Владивосток, Россия

Секция 2.3 Математическая эпидемиология

Кабанихин Сергей Игоревич, Институт вычислительной математики и математической геофизики СО РАН, Новосибирск, Россия

Романюха Алексей Алексеевич, Институт вычислительной математики им. Марчука РАН, Москва, Россия

Криворотько Ольга Игоревна, Институт вычислительной математики и математической геофизики СО РАН, Новосибирск, Россия

СИМПОЗИУМ 3. Структурная биология и фармакология: компьютерные и экспериментальные подходы

Секция 3.1. Структурная биология белков, нуклеиновых кислот и мембран

Ломзов Александр Анатольевич, Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия

Ефремов Роман Гербертович, Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН, Москва, Россия

Жарков Дмитрий Олегович, Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия

Титов Игорь Иванович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Высоцкий Евгений Степанович, Институт биофизики СО РАН, Красноярск, Россия

Иванисенко Владимир Александрович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Секция 3.2. Фармакология, хемоинформатика и химическая биология

Поройков Владимир Васильевич, Институт биомедицинской химии, Москва, Россия

Ломзов Александр Анатольевич, Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия

СИМПОЗИУМ 4. Эволюционная, популяционная и медицинская геномика/генетика человека: компьютерные и экспериментальные подходы

Секция 4.1. Популяционная и эволюционная геномика/генетика человека

Деревянко Анатолий Пантелеевич, Институт археологии и этнографии СО РАН, Новосибирск, Россия

Молодин Вячеслав Иванович, Институт археологии и этнографии СО РАН, Новосибирск, Россия

Пилипенко Александр Сергеевич, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Янковский Николай Казимирович, Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия

Хуснутдинова Эльза Камилевна, Уфимский федеральный исследовательский центр РАН, Уфа, Россия

Секция 4.2. Медицинская геномика/генетика человека

Степанов Вадим Анатольевич, Томский национальный исследовательский медицинский центр РАН, Томск, Россия

Рогаев Евгений Иванович, Университет «Сириус», Сочи, Россия

Секция 4.3. Полногеномный поиск ассоциаций

Тимощук Анна Николаевна, Московский государственный университет им. М.В. Ломоносова, Москва, Россия

СИМПОЗИУМ 5. Эволюционная, популяционная геномика/генетика и молекулярная филогения: компьютерные и экспериментальные подходы

Секция 5.1. Популяционная и эволюционная генетика/геномика диких и домашних животных

Маркель Аркадий Львович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Трапезов Олег Васильевич, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Юдин Николай Серафимович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Секция 5.2. Молекулярная филогенетика и филогеномика: растения и грибы, протисты, прокариоты и вирусы

Щербаков Дмитрий Юрьевич, Лимнологический институт СО РАН, Иркутск, Россия

Троицкий Алексей Викторович, НИИ физико-химической биологии им. А.Н. Белозерского МГУ, Москва, Россия

СИМПОЗИУМ 6. Генетика/геномика, биоинформатика и системная биология растений

Салина Елена Артемовна, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Самсонова Мария Георгиевна, Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

Афонников Дмитрий Аркадьевич, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Землянская Елена Васильевна, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия; Новосибирский государственный университет, Новосибирск, Россия

СИМПОЗИУМ 7. Генетика/геномика, биоинформатика и системная биология животных

Секция 7.1. Геномика, генетика и системная биология животных

Мошкин Михаил Павлович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Секция 7.2. Животные – генетические модели патологий человека (позвоночные и беспозвоночные)

Амстиславская Тамара Геннадьевна, НИИ нейронаук и медицины, Новосибирск, Россия

Секция 7.3. Нейрогеномика и генетика поведения

Науменко Владимир Сергеевич, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

СИМПОЗИУМ 8. Микробиология и биотехнологии: компьютерные и экспериментальные подходы

Секция 8.1. Биотехнологии через призму микробиома

Даниленко Валерий Николаевич, Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия

Припутневич Татьяна Валерьевна, Институт микробиологии, антимикробной терапии и эпидемиологии, ФГБУ НМИЦ АГП им. В.И. Кулакова Минздрава России, Москва, Россия

Секция 8.2. Микробные сообщества природных и антропогенных мест обитания

Пименов Николай Викторович, ФИЦ «Фундаментальные основы биотехнологии» РАН, Москва, Россия

Равин Николай Викторович, ФИЦ «Фундаментальные основы биотехнологии» РАН, Москва, Россия

Бонч-Осмоловская Елизавета Александровна, Московский государственный университет им. М.В. Ломоносова, Москва, Россия

Секция 8.3. Промышленные биотехнологии: создание штаммов-продуцентов

Пельтек Сергей Евгеньевич, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Брянская Алла Викторовна, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Секция 8.4. Моделирование и компьютерный анализ микробиологических систем и процессов

Акбердин Илья Ринатович, Биософт; ФИЦ Институт цитологии и генетики СО РАН, Новосибирский государственный университет, Новосибирск, Россия; Университет «Сириус», Сочи, Россия

Казанцев Федор Владимирович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

СИМПОЗИУМ 9. Биомедицина, биоинформатика и системная компьютерная биология

Секция 9.1. Экспрессия генов и болезни человека

Рагино Юлия Игоревна, НИИ терапии и профилактической медицины – филиал ФИЦ ИЦиГ СО РАН, Новосибирск, Россия

Воевода Михаил Иванович, НИИ терапии и профилактической медицины – филиал ФИЦ ИЦиГ СО РАН, Новосибирск, Россия

Орлов Юрий Львович, Первый Московский государственный медицинский университет им. И.М. Сеченова, Москва, Россия; ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Секция 9.2. Молекулярная патология, диагностика и терапия

Королев Максим Александрович, НИИ клинической и экспериментальной лимфологии – филиал ФИЦ ИЦиГ СО РАН, Новосибирск, Россия

Климонтов Вадим Валерьевич, НИИ клинической и экспериментальной лимфологии – филиал ФИЦ ИЦиГ СО РАН, Новосибирск, Россия

Коненков Владимир Иосифович, НИИ клинической и экспериментальной лимфологии – филиал ФИЦ ИЦиГ СО РАН, Новосибирск, Россия

Зельман Владимир Лазаревич, Университет Южной Калифорнии, Лос-Анджелес, США

Покровский Андрей Георгиевич, Новосибирский государственный университет, Новосибирск, Россия

Секция 9.3. Редактирование генов и геномов в моделировании заболеваний человека

Закиян Сурен Минасович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Секция 9.4. Инновационная фармакология

Салахутдинов Нариман Фаридович, Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия

Толстикова Татьяна Генриховна, Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия

Михаил Владимирович Хвостов, Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия

Павел Геннадьевич Мадонов, НИИ клинической и экспериментальной лимфологии – филиал ФИЦ ИЦиГ СО РАН, Новосибирск, Россия

Секция 9.5. Тканевая инженерия

Карпенко Андрей Анатольевич, Национальный медицинский исследовательский центр им. академика Е.Н. Мешалкина, Новосибирск, Россия

Антонова Лариса Валерьевна, НИИ комплексных проблем сердечно-сосудистых заболеваний, Кемерово, Россия

Секция 9.6. Интерстициальное пространство и длинные внесосудистые дренажно-транспортные пути

Летягин Андрей Юрьевич, НИИ клинической и экспериментальной лимфологии – филиал ФИЦ ИЦиГ СО РАН, Новосибирск, Россия

Бгатова Наталия Петровна, НИИ клинической и экспериментальной лимфологии – филиал ФИЦ ИЦиГ СО РАН, Новосибирск, Россия

СИМПОЗИУМ 10. Общие проблемы в изучении когнитивных процессов; модели когнитивной деятельности

Черниговская Татьяна Владимировна, Институт когнитивных исследований СПбГУ, Санкт-Петербург, Россия

Озгорен Мурат, кафедра биофизики, Медицинский факультет, Ближневосточный университет, Никосия, Северный Кипр

Афтанас Любомир Иванович, НИИ нейронаук и медицины, Новосибирск, Россия

Савостьянов Александр Николаевич, НИИ нейронаук и медицины, Новосибирск, Россия

Ушаков Дмитрий Викторович, Институт психологии РАН, Москва, Россия

Вальдес-Соса Педро Антонио, Университет электронных наук и технологий, Чэнду, Сычуань, Китай; Кубинский центр нейробиологии, Гавана, Куба

СИМПОЗИУМ 11. Фундаментальные генетические/клеточные системы/процессы: компьютерные и экспериментальные подходы

Секция 11.1. Репликация и репарация ДНК

Лаврик Ольга Ивановна, Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия

Секция 11.2. Транскрипция, сплайсинг, трансляция

Лаврик Ольга Ивановна, Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия

Секция 11.3. Апоптоз и другие фундаментальные клеточные процессы, регулирующие судьбу клетки

Лаврик Инна Николаевна, Университет Отто фон Герике, Магдебург, Германия

СИМПОЗИУМ 12. Математические проблемы биоинформатики и системной компьютерной биологии. Анализ больших генетических данных и искусственный интеллект

Секция 12.1. Математическое и имитационное моделирование, цифровые двойники

Голубятников Владимир Петрович, Институт математики им. Соболева СО РАН, Новосибирск, Россия

Пальянов Андрей Юрьевич, Институт систем информатики им. А.П. Ершова СО РАН, Новосибирск, Россия

Колпаков Федор Анатольевич, Университет Сириус, Сочи; ФИЦ информационных и вычислительных технологий, Новосибирск; Биософт, Новосибирск, Россия

Лашин Сергей Александрович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Марченко Михаил Александрович, Институт вычислительной математики и математической геофизики СО РАН, Новосибирск, Россия

Секция 12.2. Математическая иммунология

Бочаров Геннадий Алексеевич, Институт вычислительной математики им. Марчука РАН, Москва, Россия; Первый Московский государственный медицинский университет им. И.М. Сеченова, Москва, Россия

Акбердин Илья Ринатович, Биософт; ФИЦ Институт цитологии и генетики СО РАН, Новосибирский государственный университет, Новосибирск, Россия; Университет «Сириус», Сочи, Россия

Секция 12.3. Теория систем, анализ больших биологических данных, онтологии и искусственный интеллект

Витяев Евгений Евгеньевич, Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия

Иванисенко Владимир Александрович, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Нечесов Андрей Витальевич, Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия

Пальчунов Дмитрий Евгеньевич, Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия

Гончаров Сергей Савостьянович, Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия

СИМПОЗИУМ 13. Генетика и системная биология старения

Колосова Наталия Гориславовна, ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия

Пасюкова Елена Генриховна, Национальный исследовательский центр «Курчатовский институт», Москва, Россия

Москалёв Алексей Александрович, НИИ биологии старения Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского, Нижний Новгород, Россия

Хохлов Александр Николаевич, Московский государственный университет им. М.В. Ломоносова, Москва, Россия

Presidium of the multiconference BGRS/SB-2024 – Cochairmen:

Professor **Kolchanov Nikolay A.,** academician of the Russian Academy of Sciences, academic adviser of the Institute of Cytology and Genetics of SB RAS, head of the Department of systems biology, Russia

Professor **Kochetov Alexey V.**, academician of the Russian Academy of Sciences, Director of Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Professor **Fedoruk Mikhail P.,** academician of the Russian Academy of Sciences, Rector of Novosibirsk State University, Novosibirsk, Russia

Sokolov Igor' A., academician of the Russian Academy of Sciences, Director of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; Dean of the Faculty of Computational Mathematics and Cybernetics of Moscow State University, Moscow, Russia

Professor **Hofestädt Ralf,** University of Bielefeld, Germany (The organizer of all previous BGRS conferences since the very beginning in 1998, together with Nikolay Kolchanov)

Professor **Chen Ming,** Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou, China

Multiconference secretariat (ICG SB RAS, Novosibirsk, Russia):

Antropova Evgeniya – Scientific Secretary of the Conference

Bukharina Tatiana

Pavlova Natalia

Program committee

SYMPOSIUM 1. Genomics, transcriptomics and bioinformatics

Section 1.1. Structural and functional organization of genomes

Ershov Nikita, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Afonnikov Dmitry, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Section 1.2. Regulatory genomics

Kulakovsky Ivan, Institute of Protein Research of RAS, Pushchino; Vavilov Institute of General Genetics, Moscow; Kazan Federal University

Makeev Vsevolod, Vavilov Institute of General Genetics of RAS, Moscow Institute of Physics and Technology, Moscow, Russia

Levitsky Victor, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Merkulova Tatyana, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Section 1.3. Fundamental and Applied 3D Genomics

Graphodatsky Alexander, Institute of Molecular and Cellular Biology of SB RAS, Novosibirsk, Russia

Fishman Veniamin, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Battulin Nariman, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Section 1.4. Genome-wide transcriptomics (differential gene expression)

Oshchepkov Dmitry, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Ponomarenko Mikhail, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

SYMPOSIUM 2. Systems computational biology

Section 2.1. Reconstruction, computational analysis and modeling of gene networks and metabolic pathways

Barberis Matteo, University of Surrey, Guildford, UK
Lashin Sergey, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Govorun Vadim, Research Institute of Systems Biology and Medicine, Rospotrebnadzor, Moscow, Russia
Akberdin Ilya, Biosoft.Ru; Institute of Cytology and Genetics of SB RAS; Novosibirsk State University, Novosibirsk, Russia; Sirius University, Sochi, Russia
Ivanisenko Vladimir, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Chen Ming, Zhejiang University, Zhejiang, China
Hofestädt Ralf, University of Bielefeld, Germany

Section 2.2. Computational analysis and modeling of population, ecological and genetics systems and processes

Frisman Efim, Institute for Complex Analysis of Regional Problems of FEB RAS, Birobidzhan, Russia
Zhdanova Oksana, Institute of Automation and Control Processes of FEB RAS, Vladivostok, Russia

Section 2.3. Mathematical epidemiology

Kabanikhin Sergey, Institute of Computational Mathematics and Mathematical Geophysics of SB RAS, Novosibirsk, Russia
Romanyukha Alexey, Marchuk Institute of Numerical Mathematics of RAS, Moscow, Russia
Krivorotko Olga, Institute of Computational Mathematics and Mathematical Geophysics of SB RAS, Novosibirsk, Russia

SYMPOSIUM 3. Structural biology and pharmacology: computational and experimental approaches

Section 3.1. Structural biology of proteins, nucleic acids and membranes

Lomzov Alexander, Institute of Chemical Biology and Fundamental Medicine of SB RAS, Novosibirsk, Russia
Efremov Roman, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of RAS, Moscow, Russia
Zharkov Dmitry, Institute of Chemical Biology and Fundamental Medicine of SB RAS, Novosibirsk, Russia
Titov Igor, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Vysotski Eugene, Institute of Biophysics of SB RAS, Krasnoyarsk, Russia
Ivanisenko Vladimir, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Section 3.2. Pharmacology, cheminformatics and chemical biology

Poroikov Vladimir, Institute of Biomedical Chemistry, Moscow, Russia
Lomzov Alexander, Institute of Chemical Biology and Fundamental Medicine of SB RAS, Novosibirsk, Russia

SYMPOSIUM 4. Evolutionary, population and medical genomics/genetics of human: computational and experimental approaches

Section 4.1. Population and evolutionary human genomics/genetics

Derevianko Anatoly, Institute of Archaeology and Ethnography of SB RAS, Novosibirsk, Russia
Molodin Vyacheslav, Institute of Archaeology and Ethnography of SB RAS, Novosibirsk, Russia
Pilipenko Aleksandr, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Yankovsky Nick, Vavilov Institute of General Genetics of RAS, Moscow, Russia
Khusnutdinova Elza, Institute of Biochemistry and Genetics, Ufa Federal Research Centre of RAS, Ufa, Russia

Section 4.2. Human medical genomics/genetics

Stepanov Vadim, Tomsk National Research Medical Center of RAS, Tomsk, Russia
Rogaev Evgeny, Sirius University, Sochi, Russia

Section 4.3. Genome-wide association studies

Timoshchuk Anna, Lomonosov Moscow State University, Moscow, Russia

SYMPOSIUM 5. Evolutionary, population genomics/genetics and molecular phylogeny: computational and experimental approaches

Section 5.1. Population and evolutionary genetics/genomics of wild and domestic animals
Markel Arcady, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Trapezov Oleg, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Yudin Nikolay, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Section 5.2. Molecular phylogenetics and phylogenomics of plants, fungi, protists, prokaryotes and viruses
Sherbakov Dmitry, Limnological Institute of SB RAS, Irkutsk, Russia
Troitsky Aleksey, A.N. Belozersky Research Institute of Physico-Chemical Biology of MSU, Moscow, Russia

SYMPOSIUM 6. Genetics/genomics, bioinformatics and systems biology of plants

Salina Elena, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Samsonova Maria, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
Afonnikov Dmitry, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Zemlyanskaya Elena, Institute of Cytology and Genetics of SB RAS; Novosibirsk State University, Novosibirsk, Russia

SYMPOSIUM 7. Genetics/genomics, bioinformatics and systems biology of animals

Section 7.1. Genomics, genetics and systems biology of animals
Moshkin Mikhail, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Section 7.2. Animal genetic models of human pathologies on vertebrates and invertebrates
Amstislavskaya Tamara, Scientific-Research Institute of Neurosciences and Medicine, Novosibirsk, Russia
Section 7.3. Neurogenomics and genetics of behavior
Naumenko Vladimir, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

SYMPOSIUM 8. Microbiology and biotechnologies: computational and experimental approaches

Section 8.1. Biotechnology through the lens of the microbiome
Danilenko Valery, Vavilov Institute of General Genetics of RAS, Moscow, Russia
Priputnevich Tatiana, Institute of Microbiology, Clinical Pharmacology and Epidemiology, Moscow, Russia.
Section 8.2. Microbial communities of natural and anthropogenic habitats
Pimenov Nikolay, Federal Research Centre “Fundamentals of Biotechnology” of RAS, Moscow, Russia
Ravin Nikolai, Federal Research Centre “Fundamentals of Biotechnology” of RAS, Moscow, Russia
Bonch-Osmolovskaya Elizaveta, Lomonosov Moscow State University, Moscow, Russia
Section 8.3. Industrial biotechnology: creation of producer strains
Peltek Sergey, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Bryanskaya Alla, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Section 8.4. Modeling and computational analysis of microbiological systems and processes
Akberdin Ilya, Biosoft.Ru; Institute of Cytology and Genetics of SB RAS; Novosibirsk State University, Novosibirsk, Russia; Sirius University, Sochi, Russia
Kazantsev Fedor, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

SYMPOSIUM 9. Biomedicine, bioinformatics and systems computational biology

Section 9.1. Gene expression and human diseases

Ragino Julia, Research Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Voevoda Mikhail, Research Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Orlov Yuriy, I.M. Sechenov First Moscow State Medical University, Moscow, Russia; Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Section 9.2. pathology, diagnostics, and therapeutics

Korolev Maxim, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Klimontov Vadim, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Konenkov Vladimir, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Zelman Vladimir, University of South California, Los Angeles, USA

Pokrovskiy Andrey, Novosibirsk State University, Novosibirsk, Russia

Section 9.3. Gene and genome editing in modeling human pathological disease processes

Zakian Suren, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Section 9.4. Innovative pharmacology

Salakhutdinov Nariman, N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry of SB RAS, Novosibirsk, Russia

Tolstikova Tatiana, N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry of SB RAS, Novosibirsk, Russia

Khvostov Mikhail, N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry of SB RAS, Novosibirsk, Russia

Madonov Pavel, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Section 9.5. Tissue engineering

Karpenko Andrey, Meshalkin National Medical Research Center, Novosibirsk, Russia

Antonova Larisa, Research Institute for Complex Issues of Cardiovascular Diseases, Kemerovo, Russia

Section 9.6. Interstitial space and long extravascular drainage/transport pathways

Letyagin Andrey, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

Bgatova Natalia, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

SYMPOSIUM 10. General problems in the study of cognitive processes; models of cognitive activity

Chernigovskaya Tatyana, Institute of Cognitive Research of St. Petersburg State University, St. Petersburg, Russia

Özgören Murat, Department of Biophysics, Faculty of Medicine, Near East University, Nicosia, Northern Cyprus

Aftanas Lyubomir, Scientific-Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

Savostyanov Aleksandr, Scientific-Research Institute of Neurosciences and Medicine, Novosibirsk, Russia

Ushakov Dmitry, Institute of Psychology of RAS, Moscow, Russia

Valdés-Sosa Pedro Antonio, University of Electronic Science and Technology, Chengdu, Sichuan, China; Cuban Center for Neurobiology, Havana, Cuba

SYMPOSIUM 11. Fundamental genetic/cellular systems/processes: computational and experimental approaches

Section 11.1. DNA replication and repair

Lavrik Olga, Institute of Chemical Biology and Fundamental Medicine of SB RAS, Novosibirsk, Russia

Section 11.2. Transcription, splicing, translation

Lavrik Olga, Institute of Chemical Biology and Fundamental Medicine of SB RAS, Novosibirsk, Russia

Section 11.3. Apoptosis and other fundamental cellular processes that regulate cell fate

Lavrik Inna, Otto von Guericke University, Magdeburg, Germany

SYMPOSIUM 12. Mathematical problems of bioinformatics and systems computational biology. Big genetic data analysis and artificial intelligence

Section 12.1. Mathematical and simulation modeling, digital twins

Golubyatnikov Vladimir, Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia
Palyanov Andrey, A.P. Ershov Institute of Informatics Systems of SB RAS, Novosibirsk, Russia
Kolpakov Fedor, Sirius University, Sochi; Federal Research Center for Information and Computational Technologies, Novosibirsk; Biosoft.Ru, Novosibirsk, Russia
Lashin Sergey, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Marchenko Mikhail, Institute of Computational Mathematics and Mathematical Geophysics of SB RAS, Novosibirsk, Russia

Section 12.2. Mathematical immunology

Bocharov Gennady, Marchuk Institute of Numerical Mathematics of RAS, Moscow, Russia; I.M. Sechenov First Moscow State Medical University, Moscow, Russia
Akberdin Ilya, Biosoft.Ru; Institute of Cytology and Genetics of SB RAS; Novosibirsk State University, Novosibirsk, Russia; Sirius University, Sochi, Russia

Section 12.3. Systems theory, big biological data analysis, ontologies and artificial intelligence

Vityaev Evgeniy, Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia
Ivanisenko Vladimir, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Nechesov Andrey, Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia
Palchunov Dmitry, Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia
Goncharov Sergey, Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia

SYMPOSIUM 13. Genetics and systems biology of aging

Kolosova Nataliya, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Pasyukova Elena, National Research Center “Kurchatov Institute”, Moscow, Russia
Moskalev Alexey, Research Institute of Biology of Aging, Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia
Khokhlov Alexander, Lomonosov Moscow State University, Moscow, Russia

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук

Директор: академик РАН *Алексей Владимирович Кочетов*
Научный руководитель: академик РАН *Николай Александрович Колчанов*
Ученый секретарь: канд. биол. наук *Галина Владимировна Орлова*
Тел.: +7 (383) 363 4985, email: gorlova@bionet.nsc.ru

Институт создан в 1957 году в числе первых институтов Сибирского отделения АН СССР. В настоящее время ИЦиГ СО РАН – мультидисциплинарный, многопрофильный биологический институт, который по праву считается одним из ведущих научных учреждений биологического профиля в России. В мае 2017 года закончился второй этап реорганизации Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук.

ФИЦ ИЦиГ СО РАН включает три филиала: Сибирский научно-исследовательский институт растениеводства и селекции (СибНИИРС), Научно-исследовательский институт клинической и экспериментальной лимфологии (НИИКЭЛ), Научно-исследовательский институт терапии и профилактической медицины (НИИТПМ).

Стратегическая цель – получение новых знаний в области генетики и клеточной биологии, разработка и применение генетических технологий для решения приоритетных задач развития научно-технологического комплекса Российской Федерации.

Приоритетные задачи – получение новых фундаментальных знаний в области общей и молекулярной генетики и клеточной биологии; разработка и внедрение генетических технологий для агропромышленного комплекса, медицины и биотехнологии.

Позиционирование ИЦиГ СО РАН осуществляется по следующим направлениям: достижение результатов, обеспечивающих технологический суверенитет и конкурентные позиции Российской Федерации в стратегически важных для государства областях, включая генетические технологии для медицины, фармакологии, биотехнологической промышленности и сельского хозяйства.

Кадровый состав. На 1 января 2024 года структура ФИЦ ИЦиГ СО РАН состояла из 142 научных подразделений, в которых работали 1462 человека, в том числе 493 научных сотрудника, из них 42 % сотрудников в возрасте до 39 лет, 2 советника РАН, 8 академиков РАН, 3 члена-корреспондента РАН, 95 докторов наук, 283 кандидата наук. На 1 января 2024 года в ФИЦ ИЦиГ СО РАН обучались 81 аспирант и 34 ординатора.

Публикации. Институт активно публикуется в российских и зарубежных журналах и является в российской биологии одним из признанных лидеров. Общее количество статей в рецензируемых журналах в 2022 году составило 661. В 2018–2022 годах в международных системах цитирования публикаций Web of Science или Scopus было опубликовано 2345 статей сотрудников ИЦиГ СО РАН. ФИЦ ИЦиГ СО РАН является лидером среди НИИ и вузов РФ по количеству статей в WoS по направлению Genetics/Heredity.

Имущественный комплекс. Земельный участок площадью 35 тыс. га, закрепленный на праве постоянного пользования; 85 тыс. м^2 рабочих площадей, расположенных на территории Советского района г. Новосибирска, Барышевского сельского совета Новосибирской области, в Искитимском и Черепановском районах и в пос. Краснообск Новосибирской области.

Адрес: 630090, Россия, Новосибирск, проспект Академика Лаврентьева, 10
тел./факс: +7 (383) 363 4980/+7 (383) 333 1278
www.icgbio.ru, email: icg-adm@bionet.nsc.ru

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

Director: Full Member of the RAS *Alexey V. Kochetov*
Academic Advisor: Professor, Full Member of the RAS *Nikolay A. Kolchanov*
Academic Secretary: Candidate of Science (Biology) *Galina V. Orlova*
Phone: +7 (383) 363 4985 ext. 1336, email: gorlova@bionet.nsc.ru

The Institute was founded in 1957, among the first institutions of the Siberian Branch of the Russian Academy of Sciences. It is situated in the Novosibirsk Akademgorodok. Presently, ICG SB RAS is an interdisciplinary biological center, which ranks among the leading biological institutions in Russia. The second step of the restructuring of the Federal Research Center Institute of Cytology and Genetics was completed in May 2017. Presently, ICG includes three affiliated branches:

- Siberian Research Institute of Plant Production and Breeding (SibRIPPB). The institute is located in Krasnoobsk Village and the Novosibirsk rural area. It conducts academic, prospective, and applied studies including the collection, examination, preservation, and utilization of plant genetic resources for obtaining new biological knowledge; expansion and improvement of crop gene pools;
- Research Institute of Clinical and Experimental Lymphology (RICEL);
- Research Institute of Internal and Preventive Medicine (RIIPM).

RICEL and RIIPM are situated in the Sovetskiy and Oktyabr'skiy districts of Novosibirsk. They conduct academic, prospective, and applied studies in molecular medicine and human genetics. They also provide medical care.

The branch "Kurchatov Genomic Center", ICG SB RAS became one of the divisions of the Federal Research Center.

Tasks of ICG SB RAS: Solution of top-priority problems in the development of the Russian science and technology sector in plant genetics and breeding, animal genetics and breeding, human genetics, biotechnology, and fundamental medicine by applying methods of molecular genetics, cell biology, and computational biology.

Strategic goal: Integrated studies in plant genetics and breeding, animal genetics and breeding, human genetics, fundamental medicine, and biotechnology by applying methods of molecular genetics, cell biology, and computational biology from the generation of academic knowledge to the solution of top priority problems set by Russian agricultural, biotechnological, biomedical, and pharmaceutical industries.

Staff: As on January 1, 2024, ICG included 142 scientific units, which employed 1462 members; of them 493 researchers, 2 RAS Advisor, 8 Full Members of the RAS, 3 Corresponding Members of the RAS, 95 Doctors of Science, and 283 Candidates of Science. ICG trains 81 postgraduates and 34 ordinators.

Publications: The Institute ranks among acknowledged leaders in Russian biology. Numerous works of its researchers are published in Russian and foreign academic journals. In 2022 the overall number of publications in peer-reviewed periodicals is 661. In 2018–2022 the WoS or Scopus published 2345 articles from ICG researchers.

Journals: Institute of Cytology and Genetics is the founder of the following journals: Vavilov Journal of Genetics and Breeding, Letters to Vavilov Journal of Genetics and Breeding, Atherosclerosis, Siberian Scientific Medical Journal, and Live Science.

Auxiliaries: Core facility "Center for Genetic Resources of Laboratory Animals", which includes the unique research unit "SPF vivarium", and seven shared access centers (www.bionet.nsc.ru/uslugi/).

The Federal Research Center Institute of Cytology and Genetics is looking to cooperate with scientific and commercial enterprises.

Address: Prospekt Lavrentyeva 10, Novosibirsk, 630090 Russia
phone: +7(383) 363 4980; fax: +7(383) 333 1278
URL: www.icgbio.ru, email: icg-adm@bionet.nsc.ru

[Высокоточная амплификация]

High-fidelity amplification

BIOSAN Biolabmix®
ГРУППА КОМПАНИЙ

Высокоточные ДНК-полимеразы Biolabmix®

Высокоточные ДНК-полимеразы обеспечивают исключительную точность, скорость и производительность для всех приложений ПЦР

Данные ферменты значительнопревосходят *Pfu* и *Taq* полимеразы по следующим параметрам:

- ПРОЦЕССИВНОСТЬ
- ТОЧНОСТЬ СИНТЕЗА
- СКОРОСТЬ АМПЛИФИКАЦИИ ФРАГМЕНТОВ
- УСТОЙЧИВОСТЬ К ИНГИБИТОРАМ ПЦР

Фьюжн ДНК-полимераза

Кат № E-11001	100 е. а.
Кат № E-11005	500 е. а.

Набор для проведения ПЦР с Фьюжн ДНК-полимеразой

Кат № KH041-100	50 мкл
Кат № KH041-500	250 мкл

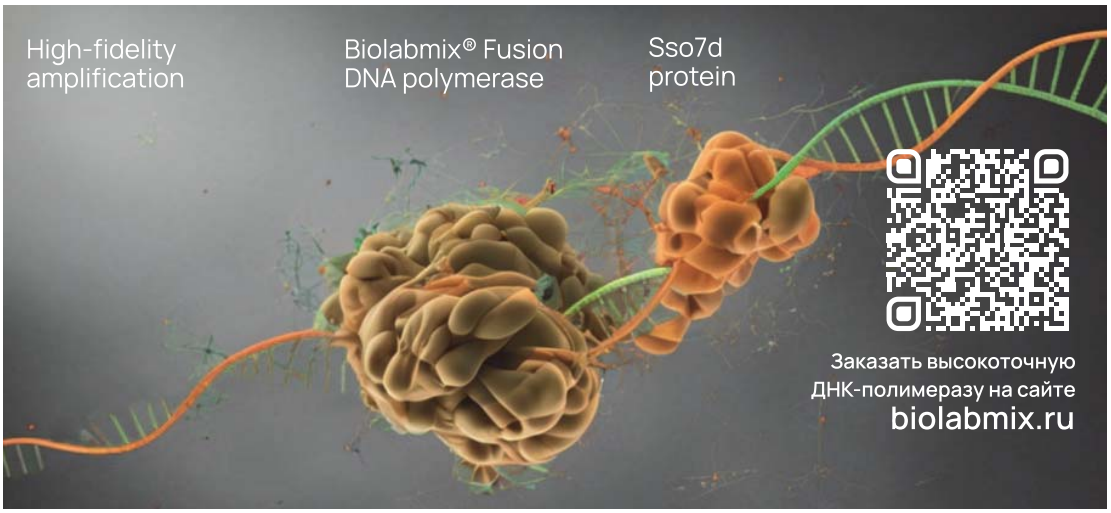
Фьюжн 2.0 полимераза

Фермент с двумя различными буферами:

- для высокоточной амплификации
- для амплификации из образцов с высоким содержанием ингибирующих веществ

Кат № E-14001	100 е. а.
Кат № E-14005	500 е. а.

Каждая партия фермента тестируется на активность фермента, электро-форетическую чистоту в SDS-ПААГ, отсутствие неспецифической нуклеаз-ной активности.



Умная линейка для выделения ДНК

D-BLOOD
Набор для выделения ДНК из образцов крови.

D-TISSUES
Набор для выделения ДНК из тканей животных и человека

D-SWABS
Набор для выделения ДНК из мазков и соскобов эпителиальных клеток, слюны

D-CELLS
Эффективный протокол выделения ДНК из клеток с отличным спектром.

D-PLANTS
Набор для выделения ДНК из растительных образцов на колонках

- Высокий выход чистой ДНК
- Широкий спектр образцов
- Выделение на колонках
- Разнообразие методов использования выделенной ДНК
- Протеиназа К в наборе для эффективного лизиса

БЕЗУПРЕЧНЫЙ СИНТЕЗ (м)РНК

**ПОЛНЫЙ СПЕКТР РЕШЕНИЙ:
ОТ RAW-MATERIALS
ДО READY-TO-USE**

АНАЛОГИ СТРУКТУРЫ КЭПА

m6AG

m7GmAmG

ARCA

МОДИФИЦИРОВАННЫЕ ТРИФОСФАТЫ

Pseudo-UTP

N1-Me-Pseudo-UTP

N6-Me-ATP

5-Me-CTP

5-OMe-UTP

ГОТОВЫЕ НАБОРЫ ДЛЯ СИНТЕЗА

Немодифицированной РНК

Кэпированной РНК

Кэпированной мРНК, содержащей в структуре модифицированные нуклеотиды: псевдоуридин (Ψ), 5-метилцитидин (m5C)

РЕАГЕНТЫ ДЛЯ СОПРЯЖЕННЫХ СТАДИЙ СИНТЕЗА (м)РНК

Наборы для мечения РНК

Высокоточные полимеразы

Наборы для выделения РНК и ДНК

BIOSAN

Biolabmix®

ГРУППА КОМПАНИЙ

8 800 600 88 76

Центрифуги для пробирок типа Eppendorf 1,5 мл
TurboFuge, Тайвань
Макс. 36 пробирок, 21 400g!

BLUE-RAY BIOTECH

Аналог MicroCL 17 (Thermo FS) и 5418R (Eppendorf)

TurboFuge Blue-Ray, Тайвань — надежные высокоскоростные центрифуги, для пробирок типа Eppendorf объемом 1,5/2,0 мл с макс. вместимостью **36** штук и ускорением **21 400g**. С использованием адаптеров возможно центрифугирование совместно с пробирками объемом 0,2 и 0,5 мл.

- Вместимость — 24 или 36 мест
- Скорость — 15 000 об/мин (21 400 g)
- Время разгона/торможения — 15/9 с
- Таймер — 0,5–99 мин

AAMC-2411	Центрифуга 15000 об/мин, 21400 g, 24×1,5 мл, крышка аэрозоль-непроницаемая метал., TurboFuge,	183 438,=
AAMC-2410	Центрифуга 15000 об/мин, 21400 g, 24×1,5 мл, крышка пластиковая, TurboFuge	183 438,=
AAMC-3611	Центрифуга 15000 об/мин, 21400 g, 36×1,5 мл, крышка аэрозоль-непроницаемая метал., TurboFuge	235 834,=
AAMC-3610	Центрифуга 15000 об/мин, 21400 g, 36×1,5 мл, крышка пластиковая, TurboFuge	227 099,=
R-2420F	Ротор угловой 24×1,5 мл	46 099,=
R-3620F	Ротор угловой 36×1,5 мл	82 978,=
AAMC-a001	Крышка ротора аэрозоль-непроницаемая, металлическая	21 205,=

Центрифуга diaGene FV2800, Россия
Сразу три ротора в комплекте!

- Максимальная вместимость — 12×1,5/2 мл (ротор R-1.5M)
- Режим вортекса для одной пробирки
- Режим работы — непрерывный и импульсный
- Габариты, Ш×Г×В — 120×170×120 мм


12×1,5 мл


12×0,5 мл + 12×0,2 мл


8 стрипов по 8×0,2 мл

FV2800	Центрифуга-вортекс, 2800 об/мин, 500g, diaGene FV2800, Диаэм, Россия	26 800,=

ДИА•М современная лаборатория

Диаэм, Москва ■ ул. Магаданская, д. 7, к. 3 ■ тел./факс: 8 (800) 234-0508 ■ sales@dia-m.ru

С.-Петербург spb@dia-m.ru	**Новосибирск** nsk@dia-m.ru	**Воронеж** vrn@dia-m.ru	**Йошкар-Ола** nba@dia-m.ru	**Красноярск** krsk@dia-m.ru
Казань kazan@dia-m.ru	**Ростов-на-Дону** rnd@dia-m.ru	**Екатеринбург** ekb@dia-m.ru	**Кемерово** kemerovo@dia-m.ru	**Нижний Новгород** nnovgorod@dia-m.ru

мобильное приложение

www.dia-m.ru

Генные технологии здоровья

СЕРИЯ G Генетический анализатор

Генетический анализатор серии G — это полностью автоматическая многоканальная платформа для генетических операций с функциями анализа фрагментов и секвенирования.
Анализатор поставляется со следующими компонентами системы: Система впрыска клея; Система оптического пути; Автоматическая система пробоотборника; Система контроля температуры; Компьютерно-программный комплекс; Модуль силовой цепи; Модуль стеллажной конструкции.

Преимущества для производительности

Широкий диапазон флуоресцентного определения

01

Твердотельный лазерный источник возбуждения, длина волны возбуждения 505 нм, диапазон дальности обнаружения до 650 нм, флуоресцентный канал обнаружения поддерживает до шести цветов.

Однородность сигнала качества

02

Режим спектральной проводимости оптического волокна, двухлучевое лазерное устройство и высокоустойчивая система оптического пути обеспечивают высокую степень однородности энергетического сигнала.

Высокая совместимость расходных материалов

03

Поддержка замены различных типов расходных материалов, таких как формамиды, полимеры и прочие ключевые расходные материалы.

Параметры	G08	G16	G24
Количество капилляров	8	16	24
Скорость обработки (36 см / режим P4)	16 образцов/час	32 образцов/час	48 образцов/час
Размеры (Г x Ш x В)	610 × 710 × 820 мм		
Масса	105 кг		
Характеристики лазера	Твердотельный лазер нового типа и высокой мощности 505 нм		
Длина волны лазера, нм	505		
Диапазон флуоресцентного определения, нм	522—650		
Мощность лазера, мВт	50		
Диапазон напряжения, кВ	0—20		
Условия эксплуатации	Влажность: 20–80 % (без конденсации); Температура в помещении: 20–30 °C, колебания температуры <±2 °C.		
Денатурация перед электрофорезом	Нет		
Уровень открытости системы	Открытая		
Время определения, мин	30		
Длина капилляров	36 см (50 см для функции секвенирования)		
Формат	.fsa/.abi		
Гарантия, лет	1		
Метод отбора проб	Автоматический отбор проб, 96-луночный планшет * 2		
Метод ввода полимера	Автоматический ввод		
Интеллектуальная подготовка образцов	Набор красителей, режим работы, приоритет и спектральная калибровка могут быть автоматически изменены группой выполнения; Поддержка приоритета автозаполнения и настройки; Поддержка переименования и запуска уже запускавшейся таблицы образцов.		
Контроль прибора	Статус планшета для образцов A/B; Контроль в реальном времени		

ДИАЭМ

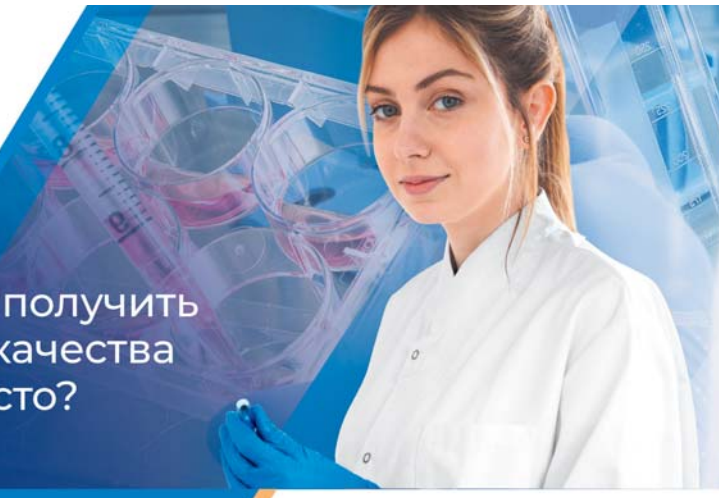
Диаэм, Москва ■ ул. Магаданская, д. 7, к. 3 ■ тел./факс: 8 (800) 234-0508 ■ sales@dia-m.ru

С.-Петербург spb@dia-m.ru
Новосибирск nsk@dia-m.ru
Воронеж vrn@dia-m.ru
Йошкар-Ола nba@dia-m.ru
Красноярск krsk@dia-m.ru
Казань kazan@dia-m.ru
Ростов-на-Дону rnd@dia-m.ru
Екатеринбург ekb@dia-m.ru
Кемерово kemerovo@dia-m.ru
Нижний Новгород nnovgorod@dia-m.ru

мобильное приложение

www.dia-m.ru

SkyGen



Хотели бы вы получить НК высокого качества быстро и просто?

kits.skygen.com

Для связи:
sales@skygen.com

SkyGen Kits NA

Широкий выбор наборов реагентов для выделения и очистки ДНК и РНК из клеток, тканей, дрожжей, нестандартных и редких образцов

Артикул: ERC423

Набор для выделения ДНК/РНК/белков
Позволяет быстро и одновременно выделять ДНК, тотальную РНК и белки из культивируемых клеток или тканей животных, а также одновременно обрабатывать большое количество различных образцов.

Весь процесс выделения может быть выполнен в течение 1 часа.

- Технология выделения: На спин-колонках
- Тип биоматериала: Ткань
- Тип биоматериала: Культура клеток
- Тип НК: ДНК/РНК/Белок
- Формат: Набор

Артикул: ERC441

Набор для выделения тотальной РНК из растительных тканей на микроцентрифужных колонках
В состав набора входит уникальный буфер СЛ, оптимально лизирующий растительные ткани. Набор проверен на таких образцах как мякоть бананов, арбузов, яблок, груш, клубни сладкого картофеля, картофеля, а также листья хлопчатника, розы, люцерны, риса и белая сосновая хвоя.

Общее время выделения составляет 1 час.

- Необходимые реагенты: β-меркаптоэтанол этанол
- Технология выделения: На спин-колонках
- Тип биоматериала: Растения
- Тип НК: РНК
- Формат: Набор

skyklad

> 4000 НАИМЕНОВАНИЙ

> 65 ПОСТАВЩИКОВ

ОТРАСЛЕВОЙ МАРКЕТПЛЕЙС

Складские товары со всей России

- Оформить заказ могут физические и юридические лица, в том числе ГОСУЧРЕЖДЕНИЯ
- Легкая оплата заказов:
  - картой онлайн
  - по счету
  - особые условия для госучреждений

АКТУАЛЬНЫЕ ЦЕНЫ

ВЕСЬ ТОВАР НА СКЛАДАХ ПОСТАВЩИКОВ В РОССИИ

БЫСТРАЯ ДОСТАВКА ОТ 3-Х ДНЕЙ

БОЛЕЕ ПОДРОБНАЯ ИНФОРМАЦИЯ

8 800 500 9442

klad.skygen.com

СИНТОЛ

НАУЧНО-ПРОИЗВОДСТВЕННАЯ КОМПАНИЯ

- Производство широкого спектра наборов реагентов и оборудования для выделения, амплификации, количественного анализа нуклеиновых кислот и их классического и массового параллельного секвенирования.
- Синтез олигонуклеотидов, зондов для ПЦР в реальном времени, модифицированных олигонуклеотидов (более 100 модификаций).
- Комплексное оснащение молекулярно-генетических лабораторий.
- Обучение молекулярно-генетическим методам исследований.
- Разработка наборов реагентов для молекулярно-генетических исследований на заказ.

127434,г. Москва, Тимирязевская, 42
syntol@syntol.ru +7(495)984-69-93
www.syntol.ru

MP Biomedicals и ее дочерняя организация в России **ООО «МПБА диагностика»** – ведущий мировой поставщик продуктов и услуг для научных исследований в области биологии и медицины. Компания предоставляет широкий ассортимент товаров, включая реагенты, лабораторное оборудование и диагностические наборы, которые используются в различных областях науки и медицины.

Почему выбирают MP Biomedicals?

1) Широкий ассортимент продукции.
   Мы предлагаем более 55000 наименований товаров, от базовых реагентов до сложных диагностических наборов и оборудования.
2) Высокое качество.
   Все наши продукты проходят строгий контроль качества и соответствуют международным стандартам.
3) Инновации.
   Наша компания постоянно вкладывает средства в исследования и разработки, предлагая рынку новейшие решения и технологии.

Основные направления деятельности:

- Молекулярная биология
  ДНК и РНК изоляция, ПЦР и qPCR реагенты, включая буферы и праймеры.
- Клеточная биология
  Продукты для культуры клеток, трансфекции и анализа белков
- Диагностика
  Наборы для диагностики инфекционных заболеваний,включая тесты на COVID-19, гепатит и ВИЧ.
- Подготовка образцов
  Высокопроизводительные системы гомогенизации, такие как FastPrep-24™ и FastPrep-96™, обеспечивающие быстрое и эффективное разрушение тканей и клеток.

Глобальное присутствие и поддержка

MP Biomedicals имеет офисы и дистрибьюторов по всему миру, включая регионы Европы, Азиатско-Тихоокеанского региона и Австралии. Это позволяет компании обеспечивать локальную поддержку и консультации для своих клиентов, предоставляя высококачественное обслуживание и экспертные знания.

История компании

Основанная более 50 лет назад, MP Biomedicals прошла путь от небольшого поставщика лабораторных реагентов до глобального лидера в области биомедицинских исследований. Сегодня компания продолжает расти и развиваться, расширяя свой ассортимент продукции и улучшая существующие технологии.

Посетите наш сайт https://www.mpbio.com/emea/ для получения дополнительной информации о продукции и услугах компании.

ООО «Компания Хеликон» – один из ведущих российских поставщиков лабораторного оборудования, реагентов и расходных материалов с 1997 года.

Компания оказывает комплекс услуг и сопровождает Клиентов на всех этапах — помогает в проектировании лабораторий, подбирает и доставляет необходимую продукцию, проводит пуско-наладку оборудования, обучает персонал на местах, обеспечивает квалифицированное сервисное обслуживание.

20 000+
наименований продукции

60+
производителей

Развитая логистическая и складская сеть

доставка в кратчайшие сроки

Направления деятельности:

- Молекулярная и клеточная биология.
- Клиническая диагностика.
- Ветеринария.
- Пищевая безопасность.
- Агрогеномика.
- Биоиндустрия.
- Криминалистика.

Для своих ключевых клиентов Компания предоставляет возможность тестирования продукции до принятия решения о покупке.

«Компания Хеликон» также имеет собственную производственную базу и выпускает лабораторное оборудование, расходные материалы и мебель под торговой маркой Helicon.

Региональные представительства Компании находятся в Санкт-Петербурге, Новосибирске, Казани, Ростове-на-Дону, Владивостоке и Екатеринбурге.

helicon
ЛУЧШИЕ РЕШЕНИЯ ДЛЯ ЛАБОРАТОРИИ

Единый телефон
8 800 770 71 21
бесплатный звонок по России

Адрес: 121374, Москва, Кутузовский проспект, д. 88
E-mail: **mail@helicon.ru**
Сайт: **www.helicon.ru**

ООО «Компания Хеликон» — один из ведущих российских поставщиков лабораторного оборудования, реагентов и расходных материалов для медицинских и прикладных задач в сфере Life Science.

Компания предлагает клиентам методическую, сервисную и техническую поддержку.

Портфолио включает более 40 мировых брендов, а также продукцию собственного производства.

Одно из ключевых преимуществ работы с компанией — возможность бесплатного тестового использования некоторых видов продукции до принятия решения о покупке. Доставка и инсталляция в лаборатории клиента осуществляется за счёт «Компании Хеликон», а развитая логистическая и складская сеть позволяет доставлять товар в кратчайшие сроки.

Контакты:

- Адрес: 121374, Москва, Кутузовский проспект, д. 88
- Телефон: +7 (800) 770 71 21 (звонки для всех абонентов на территории РФ бесплатный)
- E-mail: mail@helicon.ru
- Сайт: www.helicon.ru

NEXTGENSEQ.TECH

Исследование генетического материала
на оборудовании Illumina и PacBio согласно мировым
стандартам качества

Фото с сайта: https://www.illumina.com/systems/sequencing-platforms/novaseq-x-plus.html

NEXTGENSEQ.TECH

Решения в области науки от официального дистрибьютора CeGaT GmbH

О Нас:

ООО "Альгавитапро" является эксклюзивным партнером CeGaT GmbH в России.

О CeGaT:

CeGaT основан в 2009 году в Германии, Тюбинген. Компания специализируется в области NGS для проведения генетической диагностики, а также кастомизированных решений для науки и фарминдустрии. CeGaT предлагает широкий спектр подходов, в том числе для исследований в области геномики, транскриптомики, анализа микробиома и многих других направлений.

Профессиональная команда CeGaT поможет Вам найти оптимальное решение и достигнуть поставленных научных целей.

Что Вы получаете:

- Проведение работ по высшим мировым стандартам качества
- Секвенирование на платформах Illumina NovaSeq X Plus/6000, PacBio Sequel IIe, Oxford Nanopore
- Возможность выбирать любые параметры и режимы секвенирования
- Высокая скорость выполнения работ
- Персональное ведение каждого проекта

Наши решения в области NGS Illumina/PacBio/Nanopore

- Полногеномное секвенирование
- Полноэкзомное секвенирование
- Транскриптомный анализ
- Shotgun метагеномный анализ
- 16S (V1-V9) профилирование
- Секвенирование готовых библиотек
- Секвенирование малых РНК
- Секвенирование РНК одиночных клеток
- Секвенирование опухолевой ДНК/РНК
- Жидкостная биопсия
- Иммуногенентика
- Кастомизированные панели

Также мы предлагаем сервис по выделению нуклеиновых кислот и биоифнорматический анализ данных секвенирования.

Контактная информация:

Тел.: +7 (995) 502 43 54

Email: cegatrus@gmail.com

Email: kirill.shur@cegat.de

nextgenseq.tech
cegat.com

Общество с ограниченной ответственностью
"АЛЬГАВИТАПРО"
ИНН 7608035377
ОГРН 1187627012957

CeGaT CAP ACCREDITED COLLEGE of AMERICAN PATHOLOGISTS DAkkS Deutsche Akkreditierungsstelle D-ML-13206-01-00

Компания **«БелБиоЛаб»,** основанная в 2015, производит продукцию для ПЦР и поставляет импортное оборудование, реактивы и расходные материалы известных мировых брендов для научных институтов, производителей диагностических тестов, лабораторий различного профиля.

Продукция собственного производства:

- **Полимеразы**:
  для рутинной PCR, три вида полимеразы с **горячим стартом**.
- **Обратные транскриптазы**:
  RevM **Hot Start** ревертаза и RevM ревертаза.
- **Реактивы для ПЦР**:
  термолабильная и обычная урацил-ДНК-гликозилаза, высокоочищенная вода, дезоксинуклеозидтрифосфаты (dNTP), ингибиторы РНКаз и др.
- **Наборы и реактивы для выделения и очистки нуклеиновых кислот**:
  наборы для выделения ДНК и РНК PuriMag на магнитных частицах, магнитные частицы, протеиназа К, полиаденилат натрия Poly A, РНазин-ингибитор рибонуклеаз.
- **Реактивы для ПЦР в формате OEM и «in bulk».** Лиофилизированные и lyo-ready реактивы для ПЦР.

Поставки импортной продукции:

- **Реактивы для молекулярной биологии**:
  Termo Fisher, Qiagen, Sigma Aldrich, Promega, Jena Bioscience.
- **Химические и общелабораторные реактивы**:
  Merk, Carl Roth, PanReac Acros Organics, MPBio, AppliChem.
- **Лабораторный пластик и расходные материалы**:
  AHN, Corning, Eppendorf, Greiner Bio-one, Whatman.
- **Реактивы и лабораторные расходники по Вашему индивидуальному запросу.**

129095, г.Москва, ул.Годовикова 9, стр.13, подъезд 13.2, офис 2.4.1

www.belbiolab.ru

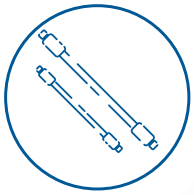
order@belbiolab.ru

+7 926 476 32 54

+7 916 120 73 36

ХИММЕД

КОМПЛЕКСНОЕ ОСНАЩЕНИЕ ЛАБОРАТОРИЙ

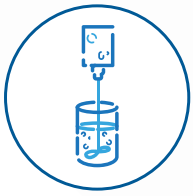


Хроматографические колонки

Стандартные образцы

Растворители для ВЭЖХ / ОСЧ

Аналитические приборы



Лабораторное оборудование

Оборудование Life Sciences

Микробиология

Химические реактивы

Биохимические реактивы

chimmed.ru

Более 20 тысяч позиций в наличии на складе в Москве!

ООО «ТД «ХИММЕД»
Москва, 115230, Каширское шоссе, дом 3, корпус 2, строение 4, этаж 6
Тел.: +7 495 640 4192, **mail@chimmed.ru**

Новосибирск, 630090, проспект Академика Лаврентьева, 6/1
Тел.: +7 383 330 9346, **sibir@chimmed.ru**

SESANA

NGS-СЕКВЕНАТОРЫ

SURFSeq 5000

FASTASeq 300

GenoLab M

БЫСТРЕЕ • ЛЕГЧЕ • ТОЧНЕЕ

• полная автоматизация приготовления NGS-библиотек •
• 96 образцов за 2 часа 30 минут •
• совместимость с реагентами Raissol Bio •

sales@sesana.ru www.sesana.ru +7 (495) 128-82-74

БИОХИМИЧЕСКИЙ БАЗАР

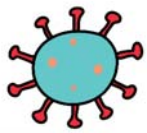
www.biohim.store

Динамично развивающаяся компания Биохимический Базар (ООО «БХБ») занимается комплексным оснащением лабораторий различного профиля.

Нашим приоритетным направлением являются лаборатории Life Sciences.

В товарную линейку входят

- **биохимические и химические реактивы:**
  - **для культур клеток**
  - **биохимии**
  - **молекулярной биологии**
  - **молекулярной генетики**
  - **микробиологии**
  - **сыворотки животных**
  - **антитела**
  - **рекомбинантные белки**
- **оборудование для лабораторий R&D**
- **расходные материалы**



Мы можем предоставить демоверсии приборов в вашу лабораторию!

Новосибирск, 630090, ул. Демакова, д. 23, офис 3.
телефоны: +8 800 700 5910, +7 903 903 4569
e-mail: bxb2021@mail.ru, mail@biohim.store

Содержание

Contents

1 Симпозиум «Геномика, транскриптомика и биоинформатика»

Symposium “Genomics, transcriptomics and bioinformatics”

doi 10.18699/bgrs2024-1.1-01

Различия в количестве генов lncRNAs вокруг гипофизарно-гонадных генов у человека и мыши

Буланов И.*, Абрамов О., Шкурат Т.
Южный федеральный университет, Ростов-на-Дону, Россия
** ibulanov@sfedu.ru*

Ключевые слова: lncRNAs; гипофизарно-гонадные гены; фолликулогенез; UCSC; NONCODE

Мотивация и цель: Репродуктивная функция человека и животных зависит от комплексных, но строго предопределенных гипофизарно-гонадных взаимодействий. Фолликулогенез – сложный и многоэтапный процесс, предшествующий овуляции. Однако начало периода полового созревания, число одновременно созревающих доминантных фолликулов, а также число детенышей в помете существенно различаются у разных видов животных. Это эволюционно закрепленные признаки, относящиеся к стойким морфофизиологическим характеристикам вида. Например, у мыши одновременно овулирует более 8 фолликулов, начиная с 2-месячного возраста животного, а у человека – один, сначала пубертатного возраста. Таким образом, гены гипофизарно-гонадной оси у разных животных начинают экспрессироваться в различное время и с существенно разным уровнем экспрессии. С другой стороны, известно, что длинные некодирующие РНК (lncRNAs), являясь жизненно важными агентами внутриклеточной генной регуляторной сети, влияют на различную жизнедеятельность клеток и экспрессию генов. LncRNA действует как молекулярная приманка, молекулярный сигнал и молекулярный каркас в организме посредством молекулярного руководства [1]. Ранее было показано, что lncRNAs регулируют развитие фолликулов и атрезию, синтез и секрецию местных гормонов яичников, а также влияют на количество фолликулов и качество ооцитов [2].
Целью данной работы было изучение особенностей структуры генов гонадотропных гормонов передней доли гипофиза – фолликулостимулирующего гормона (ФСГ) – *FSHB*, и лютеинизирующего гормона (ЛГ) – *LHB*, а также генов их рецепторов (*FSHR* и *LHCGR*) и проведение биоинформатического поиска локализации длинных некодирующих РНК в окрестностях этих генов у человека и мыши.
Методы и алгоритмы: Для изучения структуры генов и получения предшествующих и последующих последовательностей использовался UCSC Genome Browser [3] и референсные геномы hg38 и mm39. Для получения последовательностей известных генов lncRNAs использовалась база данных NONCODE [4]. Для поиска локализации полученных последовательностей lncRNAs в области 1000 нуклеотидов до гена и 1000 нуклеотидов после использовался инструмент BLAST.
Результаты: Рассмотрев гены гипофизарно-гонадной оси у человека и мыши, мы выявили практически схожие характеристики: длина гена *LHB* у человека и мыши составляет 1112 и 1078 нуклеотидов соответственно. Длина гена *FSHB* у человека и мыши – 4181 и 3517 нуклеотидов соответственно. Длина гена *LHCGR* – 68951 и

54703 нуклеотида соответственно. Длина гена *FSHR* – 192359 и 215724 нуклеотида соответственно.
Число экзонов в генах *LHB* и *FSHB* у человека и мыши – 3. Ген *LHCGR* у человека и мыши содержит по 11 экзонов, ген *FSHR* – по 10 экзонов. Процентный GC состав демонстрирует также схожие значения у человека и мыши: ген *LHB* содержит 64.12 и 59.56 % GC соответственно. Ген *FSHB* – 36.71 и 37.28 % GC соответственно. Ген *LHCGR* содержит по 40.86 и 44.42 % GC соответственно. Ген *FSHR* содержит по 39.06 и 40.33 % GC соответственно.
Несмотря на схожие структурные характеристики генов гипофизарно-гонадной оси у человека и мыши, их проксимальные окружающие последовательности (1000 нуклеотидов до гена и 1000 нуклеотидов после гена) содержат разное количество генов lncRNAs: вокруг гена *LHB* обнаруживается более чем двукратное преобладание генов днРНК у человека по сравнению с мышью (более 100 против 51 соответственно). Вокруг гена *FSHR* количество lncRNAs генов у человека превышает более чем в четыре раза количество генов lncRNAs у мыши (более 100 против 59 соответственно). Вокруг гена *FSHR* количество генов lncRNAs более чем в 1.55 раза выше, чем у мыши (92 против 59 соответственно). Вокруг гена *LHCGR* пропорция количества генов lncRNAs сохраняется – приблизительно 1:1.
Выводы: Существенные различия в количестве длинных некодирующих РНК вокруг генов гонадо-гипофизарной оси у животных с разной манифестацией периода полового созревания и числом одновременно созревающих доминантных фолликулов, свидетельствуют об их возможной важной роли в регуляции этих процессов.
Финансирование: Исследование выполнено при финансовой поддержке гранта Российского научного фонда (РНФ) № 23-15-00464.

The differences in the number of lncRNAs genes around pituitary gonadal genes in humans and mice

Bulanov I.*, Abramov O., Shkurat T.
Southern Federal University, Rostov-on-Don, Russia
** ibulanov@sfedu.ru*

Key words: lncRNAs; pituitary gonadal genes; folliculogenesis; UCSC; NONCODE

Motivation and Aim: The reproductive function of humans and animals depends on complex, but strictly predetermined pituitary-gonadal interactions. Folliculogenesis is a complex and multi-stage process preceding ovulation. However, the onset of puberty, the number of simultaneously maturing dominant follicles, as well as the number of cubs in the litter are significantly different in different animal species. These are evolutionarily fixed features related to the persistent morpho-physiological characteristics of the species. For example, in a mouse, more than 8 follicles are ovulating at the same time, starting from the age of 2 months of the animal, and in a human, one is at first puberty. Thus, the genes of the pituitary-gonadal axis in different animals begin to be expressed at different times, and with significantly different levels of expression. On the other hand, it is known that long non-coding RNAs (lncRNAs),

being vital agents of the intracellular gene regulatory network, affect various cell vital functions and gene expression. lncRNA acts as a molecular decoy, molecular signal, and molecular scaffold in the body through molecular guidance [1]. It has previously been shown that lncRNAs regulate follicle development and atresia, synthesis and secretion of local ovarian hormones, and also affect the number of follicles and the quality of oocytes [2].

The purpose of this work was to study the structural features of the genes of gonadotropic hormones of the anterior pituitary gland – follicle stimulating hormone (FSH) – *FSHB*, and luteinizing hormone (LH) – *LHB*, as well as their receptor genes (*FSHR* and *LHCGR*), and to conduct a bioinformatic search for the localization of long non-coding RNAs in the vicinity of these genes in humans and mice.

Methods and Algorithms: To study the structure of genes and obtain the preceding and subsequent sequences, the UCSC Genome Browser [3] and the reference genomes hg38 and mm39 were used. The NONCODE database [4] was used to obtain sequences of known lncRNAs genes. The BLAST tool was used to search for localization of the obtained lncRNAs sequences in the region of 1000 nucleotides before the gene and 1000 nucleotides after.

Results: After examining the genes of the pituitary-gonadal axis in humans and mice, almost similar characteristics were revealed: the length of the *LHB* gene in humans and mice is 1112 and 1078 nucleotides, respectively. The length of the *FSHB* gene in humans and mice is 4181 and 3517 nucleotides, respectively. The length of the *LHCGR* gene is 68951 and 54703 nucleotides, respectively. The length of the *FSHR* gene is 192359 and 215724 nucleotides, respectively. The number of exons in the *LHB* and *FSHB* genes in humans and mice is 3. The *LHCGR* gene in humans and mice contains 11 exons each, the *FSHR* gene contains 10 exons each. The percentage GC composition also demonstrates similar values in humans and mice: the *LHB* gene contains 64.12 and 59.56 % GC, respectively. The *FSHB* gene is 36.71 and 37.28 % GC, respectively. The *LHCGR* gene contains 40.86 and 44.42 % GC, respectively. The *FSHR* gene contains 39.06 and 40.33 % GC, respectively. Despite the similar structural characteristics of the pituitary-gonadal axis genes in humans and mice, their proximal surrounding sequences (1000 nucleotides before the gene and 1000 nucleotides after the gene) contain a different number of lncRNAs genes: around the *LHB* gene, there is more than a twofold predominance of dnRNA genes in humans than in mice (more than 100, versus 51, accordingly). Around the *FSHR* gene, the number of lncRNAs genes in humans exceeds by more than four times the number of lncRNAs genes in mice (more than 100 versus 59, respectively). Around the *FSHR* gene, the number of lncRNAs genes is more than 1.55 times that of the mouse (92 versus 59, respectively). Around the *LHCGR* gene, the proportion of the number of lncRNAs genes remains approximately 1:1.

Conclusion: Significant differences in the number of long non-coding RNAs around the genes of the gonado pituitary axis in animals with different manifestations of puberty and the number of simultaneously maturing dominant follicles indicate their possible important role in regulating these processes.

Funding: The research was carried out with the financial support of the grant of the Russian Science Foundation (RNF) No. 23-15-00464.

Список литературы/References

1. He C., Wang K., Gao Y., Wang C., Li L., Liao Y., Hu K., Liang M. Roles of noncoding RNA in reproduction. *Front Genet*. 2021;12:777510. doi 10.3389/fgene.2021.777510
2. Lv Z., Lv Z., Song L., Zhang Q., Zhu S. Role of lncRNAs in the pathogenic mechanism of human decreased ovarian reserve. *Front Genet*. 2023;14:1056061. doi 10.3389/fgene.2023.1056061
3. Nassar L.R, Barber G.P, Benet-Pagès A., Casper J., Clawson H., Diekhans M., Fischer C., Gonzalez J.N., Hinrichs A.S., Lee B.T., Lee C.M., Muthuraman P., Nguy B., Pereira T., Nejad P., Perez G., Raney B.J., Schmelter D., Speir M.L., Wick B.D., Zweig A.S., Haussler D., Kuhn R.M., Haeussler M., Kent W.J. The UCSC genome browser database: 2023 update. *Nucleic Acids Res*. 2023;51(D1):D1188-D1195. doi 10.1093/nar/gkac1072.
4. Fang S., Zhang L., Guo J., Niu Y., Wu Y., Li H., Zhao L., Li X., Teng X., Sun X., Sun L., Zhang M.Q., Chen R., Zhao Y. NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* 2018;46(D1):D308-D314. doi 10.1093/nar/gkx1107

doi 10.18699/bgrs2024-1.1-02

Гены развития и домашнего хозяйства в геноме *Drosophila melanogaster*

Жимулев И.[1]*, Ватолина Т.[1], Левицкий В.[2], Цуканов А.[2]

[1] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

** zhimulev@mcb.nsc.ru*

Ключевые слова: политенные хромосомы; гены домашнего хозяйства; гены развития; мотивы промоторов; *Drosophila*; диски и междиски; bands; interbands; белки CHRIZ/CHROMATOR

Мотивация и цель: Проблему генов «домашнего хозяйства» и генов развития мы рассматривали еще в 1975 г. Для этого использовали политенные хромосомы *Drosophila melanogaster*. Эти хромосомы постоянно находятся в состоянии интерфазы, что позволило наблюдать гены как в активном, так и в неактивном состоянии.

Политенные хромосомы имеют поперечный рисунок дисков и междисков, а также гигантские размеры и хорошо различимые два типа дисков: черные и серые. О генетической организации этих хромосомных структур было мало известно. В начале 1970-х гг. внимание авторов привлекло то обстоятельство, что для постоянной активности генов материал хромосомы также должен быть постоянно открытым для транскрипции (или деконденсированным). Этим свойством обладают только междиски, а неактивные гены должны быть расположены в плотно конденсированных дисках, поэтому все гены генома были гипотетически разделены на две группы генов: развития и домашнего хозяйства, локализованные соответственно в междисках – гены домашнего хозяйства, и в дисках – гены развития [1].

Результаты: Лишь много лет спустя мы разработали методы, позволяющие идентифицировать материал дисков и междисков хромосом на молекулярной (физической) карте ДНК, в результате чего можно определить, какие гены расположены в той или иной структуре хромосомы. Был предложен метод локализации транспозонов в политенных хромосомах под электронным микроскопом (ЭМ) и последующего выявления сайта встраивания последовательности нуклеотидов на физической карте (например, транспозон P-lArB, содержащий концевые последовательности Р-элемента, а также части lac-оперона *E. coli* и части генов дрозофилы *ADH*, *rosy* и плазмиды Bluescript). Из материала транспозона в хромосоме образовывался новый диск, что можно было увидеть под ЭМ, если внедрение происходило в междиск [2].

Затем были использованы данные, полученные в ходе реализации программы modENCODE (модельный организм ENCyclopedia of DNA Elements) [3], и, таким образом, картировали белки, расположенные в междисках близко к месту встраивания транспозона [2]. Использовали также данные трех типов классификаций хроматина в полном геноме дрозофилы [4, 6]. Для классификаций использовали модификации гистонов и белки хроматина (H3K36Me, RNApolII, Myc, инсуляторные белки BEAF-32, Mod (mdg4), Su(Hw), Zw5, CTCF, CP190),

ассоциированные с промоторами генов белки Z4 и Chriz, а также белки, локализованные в ориджинах репликации (ORC2 и ORC6). Эти и другие белки активного хроматина были использованы для построения тепловых карт колокализации данных белков [4–6]. Далее мы сосредоточились на распределении и анализе типов хроматина, которые были идентифицированы с помощью 4HMM (скрытая марковская модель). Эта модель учитывает данные связывания белков «открытого хроматина» и других белков для клеточных линий Kc, S2, BG3 и Clone8. В результате сформировали четыре основных типа хроматина, названных нами аквамарином, лазуритом, малахитом и рубином [2, 7].

Согласно этой модели и данным последующих исследований, установлен белковый состав всех четырех состояний. Состояние хроматина «аквамарин» (13 % генома), локализованного в междисках политенных хромосом, имеет наиболее обширный перечень белков и модификаций гистонов. Самым распространенным состоянием хроматина в геноме является рубин (черные диски) (48 % генома). Проверка локализации этих типов хроматина ясно показала, что аквамарин связан с междисками со значительным обогащением междискового белка CHRO (CHRIZ) [2, 72] и 5'-UTR генов. Лазуритовый хроматин соответствует телам генов и морфологически соответствовал серым дискам, соседствующим с междисками. Сделан вывод о том, что ген «домашнего хозяйства» занимает две структуры политенных хромосом: промотор находится в междиске [2, 7], а тело гена – в сером диске ниже.

В результате анализа баз данных по данным RNA-seq обнаружено, что количество транскриптов в хроматине состояния «аквамарин» в 22–27 раз превышает таковое в хроматине состояния «рубин». Путем подсчета транскриптов в различных типах клеток дрозофилы, специфичных для той или иной стадии развития, в тканях или органах личинок и взрослых мух, культурах клеток и т. п. было показано, что транскрипционно активные гены в основном расположены в междисках (аквамариновом хроматине).

Для более глубокого понимания биологических функций белков, кодируемых генами, локализующимися в четырех различных состояниях хроматина в геноме дрозофилы, мы использовали биоинформатический ресурс Gene Ontology-DAVID [8] для анализа больших чисел генов. Установлено, что гены, промоторы которых связаны с хроматином «аквамарин», участвуют в метаболических процессах, таких как протеолиз, окислительно-восстановительный процесс, в различных процессах биосинтеза ДНК, РНК, белков, липидов и пептидов и т. д. Гены развития участвуют в восприятии запаха, вкуса, химических раздражителей, транспорте ионов, развитии кутикулы и метаболическом процессе хитина, размножении многоклеточных организмов, реакции на бактерии, поведение, нейропептидный сигнальный путь, реакциях на углекислый газ, циркадный ритм и т. д.

В результате дальнейшего биоинформатического анализа мы установили, что геном содержит два типа генов: из 13 574 белок-кодирующих генов дрозофилы (FlyBase R 5.57) 6562 гена, 5'-области которых (от –300 до +200 п. н.) лежат в аквамарине, и 3162 гена развития – в рубине.

Используя инструмент STREME, мы провели поиск мотивов *de novo* в каждом из состояний хроматина и обнаружили четыре мотива, характерных только для генов домашнего хозяйства. Было обнаружено, что они содержат четыре специфических мотива, тогда как гены развития – только один TATA бокс. Около 30 % генов домашнего хозяйства не несут в себе ни одного из вышеупомянутых мотивов.

Выявлены особенности расположения генов этих групп в геноме: они различаются по длинам генов, интронов, экзонов и межгенных спейсеров, а также по количеству транскриптов, продуцируемых этими генами, и количеству экзонов в генах этих двух групп.
Гены развития имеют гораздо более длинные интроны, чем интроны генов домашнего хозяйства (в 4.5 раза больше). Гены домашнего хозяйства распределены значительно плотнее в геноме: межгенные спейсеры в пять раз короче, чем в кластерах генов развития.
Выводы: Предложена модель организации генома, позволяющая выделить и охарактеризовать две группы генов – домашнего хозяйства и развития, которые различаются по многим признакам: локализацией в структурах хромосом, различиями состояний хроматина, степенью укладки материала хромосомы и наборами белков и модификаций гистонов в местах их локаций, организацией процесса репликации (ранняя и поздняя), обилием сайтов инициации репликации (ORC). Значительно различаются структура самих генов, распределение мотивов в промоторах генов, размеры межгенных спейсеров, длина и числа интронов и экзонов, числа транскриптов с одного гена, участие в процессах инсуляции хроматина различных белков. Таким образом, фактически в пределах одного генома существуют два генома, каждый со своей организацией.
Финансирование: Работа поддержана грантом РНФ (№ 24-14-00133).

Developmental and housekeeping genes in the genome of *Drosophila melanogaster*

Zhimulev I.[1]*, Vatolina T.[1], Levitsky V.[2], Tsukanov A.[2]
[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
* *zhimulev@mcb.nsc.ru*

Key words: polytene chromosomes; housekeeping genes; developmental genes; promoter motifs; *Drosophila*; bands; interbands; CHRIZ/CHROMATOR proteins

Motivation and Aim: We considered the problem of “housekeeping” genes and developmental genes back in 1975. For this purpose we used polytene chromosomes of *Drosopila melanogaster*. These chromosomes are constantly active in interphase, which allowed us to observe genes in both active and inactive states.
Polytene chromosomes have a transverse pattern of bands and interbands, as well as giant size and well-distinguishable two types of bands: black and gray. Little was known about the genetic organization of these chromosomal structures. In the early 1970s, the authors' attention was drawn to the fact that for constant gene activity, the chromosome material must also be permanently open for transcription (or decondensed. Only interbands have this property, however inactive genes should be located in densely condensed bands, so all genes of the genome were hypothetically divided into two groups: developmental and housekeeping, localized respectively in interbands – housekeeping genes, and in bands – developmental genes [1].
Results: It was not until many years later that we developed methods to identify the bands and interbands material of chromosomes on a molecular (physical) map of DNA, which

could then determine which genes are located in a particular chromosome structure. A method was proposed to localize transposons in polytene chromosomes under an electron microscope (EM) and then identify the nucleotide sequence insertion site on the physical map (e. g., the P-lArB transposon containing the P-element end sequences, as well as parts of the lac operon of *E. coli* and parts of the *Drosophila ADH*, *rosy*, and Bluescript plasmid genes. The transposon material formed a new band in the chromosome, which can be seen under EM if the insertion took place in the interband [2].

Then, we used the data from the modENCODE program (model organism ENCyclopedia of DNA Elements) [3] and thus mapped proteins located in interbands close to the transposon insertion site [2]. We also used data from three classifications of chromatin in the full *Drosophila* genome [4, 6]. These are histone modifications H3K36Me, RNApolII, Myc, insulator proteins BEAF-32, Mod (mdg4), Su(Hw), Zw5, CTCF, CP190, gene promoter-associated proteins Z4 and Chriz, and proteins localized in replication origins (ORC2 and ORC6). These and other active chromatin proteins were placed on a heat map [4–6]. Next, we focused on the distribution and analysis of chromatin types that were identified using 4HMM (hidden Markov model). This model takes into account the binding data of "open chromatin" proteins and other proteins for Kc, S2, BG3 and Clone8 cell lines. As a result, four main types of chromatin were formed, which we named aquamarine, lazurite, malachite, and ruby [2, 7].

According to this model and the data of subsequent studies, the protein composition of all four states has been established. The aquamarine chromatin state (13 % of the genome), localized in the interbands of polytene chromosomes, has the most extensive list of proteins and histone modifications. The most common chromatin state in the genome is ruby (black disks) (48 % of the genome. Inspection of the localization of these chromatin types clearly showed that aquamarine is associated with interbands, with significant enrichment of the interband protein CHRO (CHRIZ) ([2, 7] and 5'-ends of genes; lazurite chromatin corresponds to gene bodies and morphologically corresponded to gray bands adjacent to interbands. It is concluded that the housekeeping gene occupies two polytene chromosome structures: the promoter is located in the interband [2, 7] and the gene body is located in the gray band below.

Analysis of the RNA-seq databases revealed that the number of transcripts in the aquamarine state chromatin is 22–27 times higher than that in the ruby one. By counting transcripts in different types of *Drosophila* cells specific for this or that developmental stage, in tissues or organs of larvae and adult flies, cell cultures, etc., it was shown that transcriptionally active genes are mainly located in interbands (aquamarine chromatin).

To better understand the biological functions of proteins encoded by genes localized in four different chromatin states in the *Drosophila* genome, we used the Gene Ontology-DAVID bioinformatics resource [8] for large gene number analysis. Genes whose promoters are associated with aquamarine chromatin were found to be involved in metabolic processes such as proteolysis, redox, various biosynthesis processes, DNA, RNA, proteins, lipids and peptides, etc. Developmental genes are involved in odor perception, taste, chemical stimuli, ion transport, cuticle development and chitin metabolic process, reproduction of multicellular organisms, response to bacteria, behavior, neuropeptide signaling pathway, response to carbon dioxide, circadian rhythm, etc.

Through further bioinformatics analysis, we found that the genome contains two types of genes: of the 13,574 protein-coding genes in *Drosophila* (FlyBase R 5.57),

6,562 genes whose 5'-regions (–300 to +200 bp) lie in aquamarine, and 3,162 developmental genes lie in ruby.
Using the STREME tool, we searched for de novo motifs in each of the chromatin states and found four motifs unique to housekeeping genes. Genes were found to contain four specific motifs, whereas developmental genes contain only one TATA box. About 30 % of housekeeping genes do not carry any of the above motifs.
The arrangement of genes from these groups in the genome was found to differ in the lengths of genes, introns, exons and intergenic spacers, as well as the number of transcripts produced by these genes and the number of exons in the genes of these two groups.
Developmental genes have much longer introns than the introns of housekeeping genes (4.5 times longer). Housekeeping genes are distributed much more densely in the genome: intergenic spacers are five times shorter in storage than in developmental gene clusters; their median value is 3769.5 bp in developmental stage genes and only 706 bp in housekeeping genes.
Conclusion: Genome organization model has been proposed to identify and characterize two groups of genes: housekeeping and developmental. They differ in many ways: localization in chromosome structures, differences in chromatin states, degree of stacking of chromosome material and sets of proteins and histone modifications at their localization sites, organization of the replication process (early and late), and abundance of replication initiation sites (ORCs). The structure of genes themselves, the distribution of motifs in gene promoters, the size of intergenic spacers, the length and number of introns and exons, the number of transcripts from one gene, and the participation of various proteins in chromatin insulation processes differ significantly. Thus, there are actually two genomes within one genome, each with its own organization.
Funding: The study is supported by RFS grant (No. 24-14-00133).

Список литературы/References

1. Zhimulev I.F., Belyaeva E.S. Proposals to the problem of structural and functional organization of polytene chromosomes. *Theor Appl Genet.* 1975;45(8):335-340
2. Zhimulev I., Vatolina T., Levitsky V., Tsukanov A. Developmental and housekeeping genes: Two types of genetic organization in the Drosophila genome. *Int J Mol Sci.* 2024;25(7):4068
3. The modENCODE Consortium et al. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science.* 2010;330(6012):1787-1797. doi 10.1126/science.1198374
4. Jamrich M., Greenleaf A.L., Bautz E.K. Localization of RNA polymerase in polytene chromosomes of Drosophila melanogaster. *Proc Natl Acad Sci USA.* 1977;74(5):2079-2083
5. Filion G.J. et al. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell*. 2010;143(2):212-224. doi 10.1016/j.cell.2010.09.009
6. Kharchenko P.V. et al. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature.* 2011;471(7339):480-485. doi 10.1038/nature09725
7. Zhimulev I.F. et al. Genetic organization of polytene chromosome bands and interbands in *Drosophila melanogaster*. *PLoS One.* 2014;9:e101631. doi 10.1371/journal.pone.0101631
8. Chintapalli V.R., Wang J., Dow J.A. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat. Genet.* 2007;39(6):715-720. doi 10.1038/ng2049

doi 10.18699/bgrs2024-1.1-03

Особенности лабораторного процесса и контроль качества при проведении полногеномных исследований

Криницына А., Беленикин М.
ООО «Эвоген», Москва, Россия
evogenlab@ya.ru

Ключевые слова: контроль качества, NGS, полногеномное секвенирование

Мотивация и цель: Активное развитие технологий высокопроизводительного секвенирования (NGS) в последние годы привело к экспоненциальному увеличению объемов исследований (таргетные панельные, полногеномные). Однако рост информативности лабораторных исследований неизбежно связан с ростом ошибок и отклонений на всех этапах лабораторного процесса. По литературным данным, доли ошибок и несоответствий, совершаемых при проведении лабораторной диагностики, составляют: 46–68 % на преаналитическом этапе, 7–13 % на аналитическом этапе и 18–47 % на постаналитическом этапе. В случае высокопроизводительных полногеномных генетических исследований наибольшее число отклонений наблюдается и допускается на постаналитическом этапе анализа данных. В рамках доклада проводится обсуждение: оценки структуры отклонений и несоответствий по результатам собственных и литературных данных при проведении полногеномных генетических исследований; эволюции этапов оценки лабораторных ошибок; контроля качества исследований; факторов, влияющих на качество исследований на преаналитическом и аналитическом этапах; особенностей биоинформационной обработки и интерпретации данных на постаналитическом этапе. Отдельное внимание в докладе уделяется необходимости обновления рекомендаций по интерпретации результатов высокопроизводительных исследований в части валидации генетических находок. В последнее время прослеживается смещение акцента от валидации отдельных генетических вариантов к валидации клинических кейсов, от валидации методом секвенирования по Сэнгеру к валидации биоинформационных пайплайнов. По данным нашей лаборатории, валидация ~95 % вариантов, секвенированных методом Сэнгера, технически избыточна при достаточном качестве определения («вызова») варианта по результатам биоинформационного анализа. Собственные и литературные данные свидетельствуют о возможности эволюционной трансформации сегрегационной валидации методом секвенирования по Сэнгеру в проведение сегрегационного анализа по результатам полногеномных исследований пробанда и родственников. Анализ 4000 клинических кейсов лиц с редкими/орфанными заболеваниями позволил оценить, что проведение полногеномных исследований в формате трио-WGS повышает эффективность выявления потенциально каузативных вариантов более чем на 15 % по сравнению с форматом моно-WGS. Также в докладе кратко обозревается опыт организации и проведения потоковых полногеномных исследований на базе собственной лаборатории (одной из крупнейшей в стране), анализируются факторы, которые необходимо учитывать при организации высокопроизводительной NGS лаборатории.

doi 10.18699/bgrs2024-1.1-04

Опероны О-антигенов патогенных бактерий *Herbaspirillum* sp. содержат фрагментированные гены

Кучур П.*, Комиссаров А.
Лаборатория прикладной геномики, SCAMT, Университет ИТМО, Санкт-Петербург, Россия
** chesnokova@scamt-itmo.ru*

Ключевые слова: О-антиген; оперон; *Herbaspirillum*; патоген

Мотивация и цель: Интерес к изучению оперонов О-антигенов внутри отдельного рода бактерии обусловлен двумя причинами. Первая причина связана с разнообразием образов жизни бактерий внутри рода *Herbaspirillum*: это мутуалисты, ассоциированные с растениями, условные патогены растений и патогены человека. Вторая причина заключается в стремительной модификации О-антигенов под воздействием биотических и абиотических факторов. Этим исследованием мы хотим ответить на вопрос, как меняются О-антигены на геномном уровне у бактерий разного образа жизни. Чтобы найти этот ответ, целью нашего исследования становится сравнительный анализ оперонов О-антигенов у симбиотических бактерий, отличающихся по поведению.
Методы и алгоритмы: Исходными данными являются полные геномные сборки бактерий рода *Herbaspirillum* из базы данных NCBI. Всего получено девять геномов, из которых три – геномы мутуалистических гербаспирилл, два генома оппортунистов и четыре генома патогенных штаммов гербаспирилл. Каждый геном был аннотирован Bakta. Поиск оперонов О-антигена велся по уже известным генам из путей KEGG “O-antigen repeat unit” и “O-Antigen nucleotide sugar biosynthesis”. Для каждого гена была получены профили HMM. Границы оперонов предсказывались с помощью веб-сервиса Operon-mapper. Опероном О-антигена считались опероны, включающие два и более гена из списка уже известных. Последовательности неаннотированных белков использовались для поиска сходных белков в BlastP. Готовые опероны визуализированы с помощью библиотеки DNA Features Viewer. 3D модели белков с поломками моделировались в SWISS-Model.
Результаты: Исходя из наблюдений, мы отмечаем уникальность состава оперонов О-антигена у бактерий разных видов внутри рода. Мутуалистические виды, среди которых *H. seropedicae* SmR1, *H. seropedicae* Z67 и *H. rubrisubalbicans* DSM 11543, имеют целостные опероны, без генных поломок, инсерционных элементов и иных нарушений. Аналогичные наблюдения были сделаны для оперонов О-антигенов оппортунистических видов гербаспирилл – *H. hiltneri* N3 и *H. frisingense* IAC152. В то же время у патогенных представителей рода наблюдались поломки отдельных генов в оперонах О-антигена.
В геноме первой патогенной гербаспириллы *H. frisingense* AU14559 было найдено четыре оперона, контролирующих синтез О-антигена. Два из четырех оперонов содержат *rfbBACD* гены, но очередность их расположения в оперонах разная. Один из них несет поломку гена *rfbA* и по последовательности генов похож на аналогичные *rfb* опероны оппортунистических видов гербаспирилл. Геном второй патогенной бактерии *H. seropedicae* AU14040 содержит частично утраченный ген

rfbC. В О-антигене третьей гербаспириллы *H. rubrisubalbicans* M1 обнаружена поломка гена гликозилтрансферазы. В результате этой поломки предположительно синтезирующийся им белок становится короче на 49 аминокислот. Четвертый патогенный представитель – *H. rubrisubalbicans* Os34 – содержит в одном из оперонов О-антигена сразу две поломки генов гликозилтрансфераз.
Выводы: Подводя итог сравнительного анализа оперонов О-антигена внутри рода *Herbaspirillum*, стоит отметить, что наша гипотеза об изменении оперонов О-антигенов у бактерий разного образа жизни подтверждается только частично. У мутуалистических и оппортунистических видов гербаспирилл мы не обнаружили каких-либо нарушений и аномалий генного состава оперонов соматического антигена, хотя ожидали увидеть минорные нарушения оперонов у оппортунистов. При этом гипотеза не может быть отвергнута полностью, поскольку у патогенных видов гербаспирилл был обнаружен ряд поломок генов О-антигена, потенциально приводящих к синтезу фрагментированных белков. Примечательным является также тот факт, что эти поломки преимущественно затрагивают гены гликозилтрансфераз и *rfb* генов. По-видимому, эти участники синтеза О-антигена являются наиболее восприимчивыми к атаке со стороны иммунной системы организма хозяина.
Финансирование: Исследование поддержано грантом Комитета по науке и высшей школе Санкт-Петербурга для студентов, аспирантов, молодых ученых, молодых кандидатов наук 2022 г. (ПСП № 22108).

O-antigen operons of pathogenic bacteria *Herbaspirillum* sp. contain fragmented genes

Kuchur P.*, Komissarov A.
Applied Genomics Laboratory, SCAMT Institute, ITMO University, St. Petersburg, Russia
** chesnokova@scamt-itmo.ru*

Key words: O-antigen; operon; *Herbaspirillum*; pathogen

Motivation and Aim: The interest in studying O-antigen operons within a single bacterial genus is motivated by two reasons. One reason is the diversity of bacterial lifestyles within the genus *Herbaspirillum*, namely the plant-associated mutualists, opportunistic plant pathogens, and human pathogens. The second reason is the rapid modification of O-antigens under biotic and abiotic selective pressures. With this study, we aim to address the question of how the genomic diversity of O-antigen operons correlates with the diverse lifestyles of the bacteria. To this end, we conduct a comparative analysis of O-antigen operons among representative symbiotic bacteria.
Methods and Algorithms: The input data for this study are the complete genome assemblies of *Herbaspirillum* strains retrieved from the NCBI database. In total, we collected nine genomes including three genomes of mutualistic *Herbaspirillum* strains, two genomes of opportunistic *Herbaspirillum* strains, and four genomes of pathogenic *Herbaspirillum* strains. Each genome was annotated using the Bakta pipeline. O-antigen operons were identified by searching for the KEGG pathways “O-antigen repeat unit” and “O-Antigen nucleotide sugar biosynthesis”. HMM profiles were obtained for each KEGG gene. Operon boundaries were predicted using the Operon-mapper web service.

O-antigen operons were defined as operons including two or more KEGG genes. The sequences of unannotated proteins were used to search for homologous proteins using BlastP. Assembled operons were visualized using the DNA Features Viewer library. 3D protein models with template breaks were modeled using SWISS-Model.

Results: Based on our observations, we note the uniqueness of the composition of the O-antigen operons in bacteria of different species within the genus. The mutualistic species, including *H. seropedicae* SmR1, *H. seropedicae* Z67, and *H. rubrisubalbicans* DSM 11543, have intact operons without gene breaks, insertion elements, or other disruptions. Similar observations were made for the O-antigen operons of the opportunistic *Herbaspirillum* species *H. hiltneri* N3 and *H. frisingense* IAC152. In contrast, disruptions of individual genes within the O-antigen operons were observed in the pathogenic representatives of the genus.

In the genome of the first pathogenic *Herbaspirillum* strain *H. frisingense* AU14559, we identified four operons controlling the synthesis of O-antigen. Two of the four operons contain the *rfbBACD* genes, but the order of their arrangement in the operons is different. One of them carries a disruption in the *rfbA* gene and resembles the *rfb* operons of the opportunistic *Herbaspirillum* species in terms of gene sequence. The genome of the second pathogenic bacterium, *H. seropedicae* AU14040, contains a partially truncated *rfbC* gene. In the O-antigen of the third *Herbaspirillum* strain, *H. rubrisubalbicans* M1, a disruption in the glycosyltransferase gene was identified. As a result of this disruption, the protein synthesized by this gene is predicted to be 49 amino acids shorter. The fourth pathogenic representative, *H. rubrisubalbicans* Os34, contains two disruptions in glycosyltransferase genes within one of the O-antigen operons.

Conclusion: Summarizing the comparative analysis of O-antigen operons within the genus *Herbaspirillum*, it is worth noting that our hypothesis about the alteration of O-antigen operons in bacteria with different lifestyles is only partially confirmed. In mutualistic and opportunistic *Herbaspirillum* species, we did not identify any disruptions or abnormalities in the gene composition of the somatic antigen operons, although we expected to find minor disruptions in the operons of the opportunists. However, the hypothesis cannot be completely rejected, since a number of disruptions in O-antigen genes were found in pathogenic *Herbaspirillum* species, potentially leading to the synthesis of truncated proteins. It is also noteworthy that these disruptions predominantly affect the glycosyltransferase and *rfb* genes. Apparently, these participants in O-antigen synthesis are the most susceptible to attack by the host immune system.

Funding: This study was supported by a grant from the Committee for Science and Higher Education of St. Petersburg for students, postgraduates, young scientists, and young candidates of sciences in 2022 (PSP No. 22108).

Список литературы/References

1. Moran N.A., Wernegreen J.J. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol*. 2000;15(8):321-326. doi 10.1016/S0169-5347(00)01902-9
2. Lerouge I., Vanderleyden J. O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions. *FEMS Microbiol Rev*. 2002;26(1):17-47. doi 10.1111/j.1574-6976.2002.tb00597.x
3. Kalynych S., Morona R., Cygler M. Progress in understanding the assembly process of bacterial O-antigen. *FEMS Microbiol Rev*. 2014;38(5):1048-1065. doi 10.1111/1574-6976.12070

doi 10.18699/bgrs2024-1.1-05

Альтернативные структуры ДНК обнаружены вблизи ультраконсервативных некодирующих элементов генома

Макарова И.*, Калюжный Д.
Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия
** irina.vic.makarova@gmail.com*

Ключевые слова: ультраконсервативные некодирующие элементы генома (УНЭ); альтернативные структуры ДНК; ДНК квадруплексы; ДНК триплексы

Мотивация и цель: Некодирующие, но при этом консервативные элементы генома сохранились практически неизменными от курицы и других эволюционно более удаленных позвоночных до человека. Причины такой эволюционной стабильности исследуют с точки зрения их нуклеотидного состава, эпигенетической регуляции, ассоциированности с сайтами связывания транскрипционных факторов и границами топологически ассоциированных доменов [1, 2]. Наша цель – расширить анализ и определить возможную роль примыкающих участков в эволюционной консервативности данных элементов.
Методы и алгоритмы: Мы проанализировали ультраконсервативные некодирующие элементы (УНЭ) в трех организмах – человеке (*Homo sapiens*), мыши (*Mus musculus*) и рыбке Данио (*Danio rerio*). Положения УНЭ были использованы из UCNEbase [3]. Положения последовательностей, способных формировать структуры, отличные от B-формы, были получены из следующих источников:

- Для квадруплексных участков использовались данные эксперимента G4-seq [4].
- Для предсказания мотивов ДНК триплексов (H-мотивов) и крестообразных структур применялся Inverted Repeats Finder версии 3.08 [5], где они были определены как зеркальные полипиримидиновые/ полипуриновые и инвертированные повторы соответственно.

Статистический анализ проводился относительно перемешанных положений структур, отличных от B-формы. Для предобработки и анализа данных использовались скрипты, написанные на языке Python версии 3.11 в редакторе кода Visual Studio Code версии 1.87.0. Анализ обогащения по функциональной принадлежности был выполнен с помощью STRING-db версии 12.0.
Результаты: При анализе нуклеотидного состава участков, примыкающих к ультраконсервативным элементам, мы обнаружили обогащение гуанином и цитозином участков, смежных с УНЭ во всех трех геномах. Мы также заметили асимметрию обогащения гуанином с 3’ конца элементов. Наиболее выраженное обогащение было отмечено для рыбки Данио.
Отмеченная неравномерность нуклеотидного состава подтолкнула нас проанализировать структуры, отличные от B-формы, специфичные к геномной последовательности, которые могли бы образовываться на участках внутри и в непосредственной близости к УНЭ.
Мы обнаружили, что участки, примыкающие к ультраконсервативным элементам с 3’ конца, значимо обогащены квадруплексами в геномах человека и мыши. В

геноме рыбки Данио достоверное обогащение характерно для примыкающих участков в обоих направлениях, при этом 3’ конец обогащен в большей степени. H-мотивы также значимо обогащали участки, прилегающие к ультраконсервативным элементам, в геномах всех трех организмов. Для H-мотивов не характерна асимметричность обогащения с одного конца, но при этом они преимущественно локализованы на кодирующий цепи ДНК. Интересно отметить, что разные ультраконсервативные элементы фланкированы квадруплексами и H-мотивами.
Анализ мотивов крестообразных структур не показал наличия паттерна ассоциированности с ультраконсервативным элементами.
Выводы: Мы впервые провели анализ ассоциации ультраконсервативных некодирующих элементов с альтернативными структурами ДНК. Значимое обогащение квадруплексными структурами и триплексными мотивами может являться ключом к пониманию устойчивости нуклеотидных последовательностей в эволюции. Наше исследование открывает перспективы изучения функциональной значимости и эволюционной стабильности ультраконсервативных некодирующих элементов.

Non-B DNA structures adjacent to ultraconserved non-coding elements unveiled

Makarova I.*, Kaluzhny D.
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
** irina.vic.makarova@gmail.com*

Key words: ultraconserved non-coding elements (UCNEs); non-B DNA structures; DNA quadruplexes; DNA triplexes

Motivation and Aim: Non-coding, yet highly conserved genomic elements persist virtually unchanged from chickens and other evolutionarily distant vertebrates to humans. The reasons for such evolutionary stability are explored in terms of their nucleotide composition, epigenetic regulation, association with transcription factor binding sites and boundaries of topologically associated domains [1, 2]. Our aim is to expand the analysis and determine the potential role of adjacent regions in the evolutionary conservation of these elements.
Methods and Algorithms: We analyzed ultraconserved non-coding elements (UCNEs) in three organisms – human (*Homo sapiens*), mouse (*Mus musculus*), and zebrafish (*Danio rerio*). Positions of UCNEs were retrieved from UCNEbase [3]. Positions of sequences capable of forming non-B DNA structures were obtained from the following sources:

- For quadruplex regions, data from G4-seq experiment [4] were utilized.
- For predicting regions of DNA triplexes (H-motifs) and cruciform structures Inverted Repeats Finder version 3.08 [5] was used. H-motifs and cruciform motifs were defined as mirror polypyrimidine/polypurine repeats and inverted repeats, respectively.

Statistical analysis was carried out regarding the shuffled positions of non-B DNA structures. Data preprocessing and analysis were conducted using Python 3.11 scripts in

the Visual Studio Code editor. Gene-set enrichment analysis was performed using STRING-db version 12.0.

Results: In all three genomes, we observed enrichment of guanine and cytosine in regions adjacent to UCNEs. We also noted an asymmetry in guanine enrichment at the 3' end of the elements, with the most pronounced enrichment observed in zebrafish.

The observed unevenness in nucleotide composition prompted us to examine non-B DNA structures that are specific to the genomic sequence and could form within or in close proximity to UCNEs. We found significant enrichment of quadruplexes in regions adjacent to ultraconserved elements at the 3' end in the human and mouse genomes. In the zebrafish genome, enrichment was observed in both directions, with a greater extent at the 3' end.

H-motifs were also significantly enriched in regions adjacent to UCNEs in all three organisms. Unlike quadruplexes, H-motifs did not exhibit enrichment asymmetry from one end but were predominantly localized on the coding DNA strand. Interestingly, different ultraconserved elements were flanked by quadruplexes and H-motifs.

Analysis of cruciform structure motifs did not reveal a pattern of association with ultraconserved elements.

Conclusion: We conducted, for the first time, an analysis of the association between ultraconserved non-coding elements and non-B DNA structures. Significant enrichment with quadruplex structures and triplex motifs may be key to understanding the stability of nucleotide sequences in evolution. Our study opens up perspectives for investigating the functional significance of ultraconserved non-coding elements.

Список литературы/References

1. Polychronopoulos D., King J.W.D., Nash A.J., Tan G., Lenhard B. Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic Acids Res*. 2017;45(22):12611-12624. doi 10.1093/nar/gkx1074
2. Fedorova L., Crossley E.R., Mulyar O.A., Qiu S., Freeman R., Fedorov A. Profound non-randomness in dinucleotide arrangements within ultra-conserved non-coding elements and the human genome. *Bology (Basel)*. 2023;12(8):1125. doi 10.3390/biology12081125
3. Dimitrieva S., Bucher P. UCNEbase-A database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic Acids Res*. 2013;41:D101-D109. doi 10.1093/nar/gks1092
4. Marsico G., Chambers V.S., Sahakyan A.B., McCauley P., Boutell J.M., Antonio M.D., Balasubramanian S. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res*. 2019;47(8):3862-3874. doi 10.1093/nar/gkz179
5. Warburton P.E., Giordano J., Cheung F., Gelfand Y., Benson G. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res.* 2004;14(10A):1861-1869. doi 10.1101/gr.2542904

doi 10.18699/bgrs2024-1.1-06

Веб-приложение speCOIdent для анализа результатов генотипирования криптических видов байкальских амфипод рода *Eulimnogammarus*

Мутин А.Д.*, Саранчина А.Е., Ржечицкий Я.А., Помазкин В.К., Тимофеев М.А., Дроздова П.Б.

Научно-исследовательский институт биологии, Иркутский государственный университет, Иркутск, Россия

** andreimutin97@gmail.com*

Ключевые слова: Байкал; амфиподы; молекулярная филогенетика; speCOIdent; цитохром-с-оксидаза (субъединица I)

Мотивация и цель: Озеро Байкал – одно из древнейших озер на планете, примечательной особенностью которого является обширное биоразнообразие фауны, насчитывающей более 2500 видов [1]. Наиболее значимым и многочисленным отрядом байкальских беспозвоночных являются амфиподы. В Байкале амфиподы представлены более чем 350 морфологическими видами и подвидами, среди которых отмечен 99 % эндемизм, однако степень криптической изменчивости остается неясной. Недавние исследования продемонстрировали, что одни из наиболее распространенных обитателей литоральной зоны озера, *Eulimnogammarus verrucosus* (Gerstfeldt, 1858) и *Eulimnogammarus vitatus* (Dybowsky, 1874), включают в себя разные генетические линии [2]. Помимо этого, у *E. verrucosus* было обнаружен репродуктивный барьер между двумя генетическими линиями данного вида, и подобные исследования продолжаются [3]. Наличие криптической изменчивости показывает, что для эффективного изучения видового разнообразия амфипод, помимо анализа морфологических различий, требуется исследование различий на генетическом уровне – генотипирование, которое обычно проводят с помощью секвенирования по Сэнгеру. Популярным инструментом для интерпретации результатов генотипирования является поиск в базе данных NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) и Barcode of Life Database (BOLD; https://boldsystems.org). Однако использовать эти сервисы для идентификации генетической группы не всегда эффективно. Пользователи NCBI BLAST часто вынуждены читать соответствующую публикацию для уточнения генетической линии, поскольку в базе Genbank могут быть указаны только виды. BOLD, в свою очередь, позволяет определять генетические группы, однако идентифицирует их как BIN (Barcode Index Number), что также требует от пользователя ознакомится с соответствующей публикацией для того чтобы соотнести какой генетической группе соответствует конкретный BIN. Помимо этого, NCBI позволяет проводить идентификацию только на основе уже опубликованных последовательностей, что усложняет рутинную работу с новыми, неопубликованными последовательностями.

Таким образом, в связи с вышеперечисленным, целью данной работы является создание удобного, веб-приложения для эффективного определения видовой

принадлежности и/или генетической линии амфипод озера Байкал путем анализа результатов генотипирования.
Методы и алгоритмы: Используемая в приложении база данных была создана с помощью языка программирования Bash на основе стандартных филогенетических маркеров (цитохром-с-оксидаза I (COI) и 18S рРНК) амфипод Байкальского региона. База данных была сформирована из новых последовательностей, полученных с помощью секвенировния по Сэнгеру и дополнена последовательностями доступными в Genbank. Веб-приложение было создано на основе языка программирования R и с помощью ключевых пакетов Shiny [4], rBLAST [5], ggplot2 [6], mapview [7], leaflet [8].
Результаты: Было проведено генотипирование некоторых видов рода *Eulimnogammarus* из озера Байкал и реки Ангара и проанализирована на наличие различных генетических групп внутри каждого из видов. На основе этих данных была создана база данных, используемая в приложении speCOIdent.
Созданное веб-приложение speCOIdent размещено на одной странице и имеет три интерактивные панели для взаимодействия с пользователем: 1) Панель для ввода последовательности и настройки параметров идентификации. Пользователь имеет возможность скопировать последовательность в окно для ввода как в формате fasta, так и в формате нуклеотидной последовательности без заголовка. Помимо этого, в приложении присутствует возможность изменять параметры идентификации, такие как «e-value» (мера вероятности того, что найденное совпадение случайно) и «% ID» (процент идентичности последовательностей). 2) Результатом идентификации является всплывающее интерактивное окно, в котором содержится информация о типе гена (COI или 18S рРНК), видовой принадлежности и генетической группе. Помимо этого, на втором окне отображается динамически масштабируемая, интерактивная карта ареала вида или генетической группы, позволяющая с высокой точностью определить известные места их обитания. 3) Также дополнительным результатом идентификации является таблица из пяти лучших совпадений, содержащая название вида, название его генетической группы, название гена, а также процент идентичности. Данная таблица является интерактивной и позволяет вывести на экран полный список параметров идентификации и выравнивание на последовательность интереса. Дополнительными функциями speCOIdent является встроенная инструкция по использованию и возможность скачать полную базу данных в формате fasta.
Помимо этого, с помощью веб-приложения speCOIdent были проанализированы последовательности COI нескольких популяций амфипод вида *E. verrucosus* из озера Байкал и реки Ангара и определены их генетические линии.
Выводы: Было создано веб-приложение speCOIdent с базой данных из последовательностей стандартных филогенетических маркеров COI и 18S рРНК для определения видовой принадлежности, генетической группы и ареала амфипод Байкальского региона путем анализа результатов генотипирования. Приложение speCOIdent сочетает в себе преимущества как NCBI BLAST, так и BOLD, позволяя быстро и с высокой точностью идентифицировать не только видовую принадлежность, но и генетическую группу, параллельно отражая маштабируемую карту ареала. Помимо этого, speCOIdent является гибкой и легко адаптируемой программой, которую можно быстро настроить на использование различных баз данных с любыми маркерными последовательностями. Данное

приложение было сделано общедоступным и размещено на сервере научно-исследовательского института биологии ИГУ по адресу (http://bioinformatics.isu.ru:3838/speCOIdent/). Помимо этого, исходный код speCOIdent был опубликован на веб-сервисе GitHub по адресу (https://github.com/MutinAndrei/Shiny-web-application-for-analyzing-genotyping-results).
Финансирование: Исследование было проведено при поддержке гранта министерства науки и высшего образования России № FZZE-2024-0008.

Web application speCOIdent for the analysis of genotyping results for cryptic species of Baikal amphipods belonging to the genus *Eulimnogammarus*

Mutin A.D.*, Saranchina A.E., Rzhechitskiy Y.A., Pomazkin V.K., Timofeyev M.A., Drozdova P.B.

Institute of Biology at Irkutsk State University, Irkutsk, Russia
** andreimutin97@gmail.com*

Key words: Baikal; amphipods; molecular phylogenetics; speCOIdent; cytochrome c-oxidase (subunit I)

Motivation and Aim: Lake Baikal is one of the oldest lakes in the world, a remarkable feature of which is the rich biodiversity of its fauna, including over 2500 species [1]. Amphipods are one of the most species-rich and important groups of Baikal invertebrates. In Baikal, amphipods are represented by more than 350 morphological species and subspecies, over 99 % of which are endemic, but the degree of cryptic diversity remains unclear. Recent studies have shown that some of the most common inhabitants of the littoral zone of the lake, *Eulimnogammarus verrucosus* (Gerstfeldt, 1858) and *Eulimnogammarus vitatus* (Dybowsky, 1874), comprise different genetic lineages [2]. In addition, a reproductive barrier between two genetic lineages of *E. verrucosus* has been found, and there is more to learn [3]. The presence of cryptic diversity shows that to effectively study amphipod species diversity, it is necessary, in addition to analyzing morphological differences, to study differences at the genetic level by genotyping, which is usually performed using Sanger sequencing. Popular tools for interpreting genotyping results are search engines of the NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and the Barcode of Life (BOLD; https://boldsystems.org) databases. However, using these services to identify a genetic group is not always efficient. NCBI BLAST users often need to read the relevant publication to clarify the genetic lineage, as the Genbank database may only list species. BOLD, on the other hand, provides the identification of genetic groups, but identifies them as BINs (Barcode Index Numbers), which also requires the user to consult the relevant publication to determine which genetic group a particular BIN corresponds to. In addition, NCBI only allows identification based on previously published sequences, which makes routine work with new, unpublished sequences difficult.
In view of the aforementioned considerations, the objective of this study is to create a user-friendly, web-based application for the efficient determination of the species and/or

genetic lineage of amphipods of Lake Baikal by means of the analysis of genotyping results.

Methods and Algorithms: The database utilized in this application was constructed using the Bash programming language, based on standard phylogenetic markers (cytochrome c-oxidase I (COI) and 18S rRNA) of amphipods from the Baikal region. The database was constructed from newly generated sequences obtained through Sanger sequencing and augmented with sequences available in GenBank. The web application was developed using the R programming language and several key packages, including Shiny [4], rBLAST [5], ggplot2 [6], mapview [7], and leaflet [8].

Results: Genotyping of some species of the genus *Eulimnogammarus* from Lake Baikal and the Angara River was conducted and analyzed to identify the presence of different genetic groups within each species. The resulting data were used to create a database, which was subsequently incorporated into the speCOIdent application.

The speCOIdent web application is comprised of a single page and features three interactive panels for user interaction: 1) a panel for entering the sequence and setting the identification parameters. The user may input the sequence in either FASTA format or nucleotide sequence format, without a header. Furthermore, the application enables the user to modify identification parameters, such as the e-value (a measure of the probability that the match is random) and the % ID (percentage of sequence identity). 2) The identification result is presented in an interactive pop-up window, which provides information about the type of the genetic marker (COI or 18S rRNA), species affiliation, and genetic group. Additionally, a second window displays a scalable, interactive map of the known records for the species or genetic group, allowing for precise identification of its range. Additionally, the identification process yields a table of the top five matches, which includes the species name, the name of the genetic group, gene name, and identity percentage. This table is interactive, allowing the user to view the full list of identification parameters and alignment to the sequence of interest. SpeCOIdent also features a built-in instruction manual and the possibility to download the full database in FASTA format.

Using the speCOIdent web application, the COI sequences of several populations of amphipod species *E. verrucosus* from Lake Baikal and the Angara River were analyzed and their genetic lineages were determined.

Conclusion: A web application, speCOIdent, was created with a database of sequences of standard phylogenetic markers, COI and 18S rRNA. The application was used to determine the species, genetic group, and range of amphipods in the Baikal region by analyzing genotyping results. The speCOIdent application integrates the strengths of both NCBI BLAST and BOLD, enabling rapid and precise identification of species affiliation and genetic group, while reflecting a user-friendly range map. Moreover, speCOIdent is a versatile and adaptable program that can be rapidly configured to utilize disparate databases with any marker sequences. The application is now publicly available and is hosted on the server of the Institute of Biology at Irkutsk State University at http://bioinformatics.isu.ru:3838/speCOIdent/. Furthermore, the source code of speCOIdent has been made publicly available on the GitHub web service at (https://github.com/MutinAndrei/Shiny-web-application-for-analyzing-genotyping-results).

Funding: The study was financially supported by grant Ministry of Science and Higher Education of Russia No. FZZE-2024-0008.

Список литературы/ References

1. Timoshkin O.A. Lake Baikal: diversity of fauna, problems of its immiscibility and origin, ecology and “exotic” communities. In: Index of animal species inhabiting Lake Baikal and its catchment area. Vol. I. Lake Baikal. Novosibirsk: Nauka, 2001;74-113
2. Gurkov A. et al. Indication of ongoing amphipod speciation in Lake Baikal by genetic structures within endemic species. *BMC Evol Biol*. 2019;19:1-16
3. Drozdova P. et al. Experimental crossing confirms reproductive isolation between cryptic species within Eulimnogammarus verrucosus (Crustacea: Amphipoda) from Lake Baikal. *Int J Mol Sci*. 2022;23(18):10858
4. Chang W. et al. shiny: Web application framework for R. 2023. https://CRAN.R-project.org/package=shiny
5. Hahsler M., Nagar A. rBLAST: R interface for the basic local alignment search tool. 2019. https://github.com/mhahsler/rBLAST
6. Wickham H. ggplot2: Elegant graphics for data analysis. Berlin: Springer, 2016
7. Appelhans T., Detsch F., Reudenbach C., Woellauer S. mapview: Interactive viewing of spatial data in R. 2023. https://cran.rproject.org/web/packages/mapview/index.html
8. Cheng J., Schloerke B., Karambelkar B., Xie Y. leaflet: Create interactive web maps with the javaScript “Leaflet” library. 2023. https://cran.rproject.org/web/packages/leaflet/index

doi 10.18699/bgrs2024-1.1-07

Актуальность интеграции данных цитогенетики и сборок геномов хромосомного уровня на примере различных групп позвоночных

Романенко С.[1]*, Беклемишева В.[1], Перельман П.[1], Андреюшкова Д.[1], Марченко С.[2], Лемская Н.[1], Билтуева Л.[1], Кливер С.[3], Трифонов В.[1], Графодатский А.[1]

[1] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Независимый исследователь, Копенгаген, Дания*

* *rosa@mcb.nsc.ru*

Ключевые слова: дифференциальная окраска хромосом; гетерохроматин; кариотип; скаффолд; флуоресцентная гибридизация *in situ*; хромосомные перестройки; центромера

Мотивация и цель: Сборки до уровня хромосом являются золотым стандартом для секвенирования всего генома *de novo* [1–3]. Получение высококачественных эталонных геномов позвоночных, собранных до хромосомного уровня, привело к значительным результатам в функциональной, сравнительной и популяционной геномике [4, 5]. Сложными этапами при построении эталонного референсного генома являются правильная ориентация контигов и скаффолдов по отношению друг к другу и сборка суперскаффолдов в структурированные хромосомные скаффолды, а также секвенирование и позиционирование участков генома с повторяющимися последовательностями или с сегментными дупликациями [6–8]. Многие из этих проблем преодолеваются при интеграции данных секвенирования генома и данных цитогенетики. Комбинированный подход, обеспечивающий сборку генома до уровня хромосом, гораздо более успешен в ответе на фундаментальные биологические вопросы, чем одни лишь геномные или цитогенетические подходы [9–11]. Зачастую сопоставление данных секвенирования генома и цитогенетических данных не проводится, что отражается в несогласованности нумерации хромосом даже для видов, для которых имеются стандартизированные кариотипы, к неверной ориентации скаффолдов, к недооценке размеров хромосом из-за неучетов блоков гетерохроматина и т. д. Целью данной работы является привлечение внимания к важности понимания и применения комбинированного подхода при выполнении сборок геномов до хромосомного уровня.

Методы и алгоритмы: В работе использовались методы классической, дифференциальной и молекулярной цитогенетики [12–15]. Микродиссекция проводилась для получения хромосомо- и районоспецифичных библиотек [16]. Приготовление библиотек для секвенирования выполнялось, как описано ранее [17]. Для анализа данных секвенирования применялись различные подходы [17–20].

Результаты: Путем парного выравнивания мы идентифицировали и визуализировали синтенные блоки, общие для морской свинки (*Cavia porcellus*), голого землекопа (*Heterocephalus glaber*) и человека. Высокое качество сборки генома человека позволило определить положение центромер при сравнении синтений.

Мы объединили мелкие синтении в блоки размером не менее 5 м.п.о. для сопоставления данных сравнительной цитогенетики и сборок геномов, полученных методом анализа данных секвенирования библиотек Hi-C, что позволило привязать геномные скаффолды к отдельным хромосомам и ориентировать их относительно центромер и блоков гетерохроматина. Данные Hi-C позволили закрыть все пробелы на сравнительных картах и выявить дополнительные перестройки, отличающие кариотипы трех видов. В результате мы интегрировали биоинформатические и цитогенетические данные и скорректировали предыдущие сравнительные карты и сборки генома морской свинки, голого землекопа и человека [19].

В кариотипе стерляди (*Acipenser ruthenus*, ARUT) из 120 хромосом цитогенетически легко идентифицируются первые 14–17 пар и пары мелких хромосом, несущие гены ЯОР [21]. Методом микродиссекции мы получили библиотеки отдельных хромосом стерляди. Эти библиотеки были помечены и использованы в качестве зондов для FISH, а также секвенированы и выровнены на сборку генома стерляди хромосомного уровня (GCF_902713425.1/), чтобы увязать цитогенетические и биоинформатические данные. Анализ результатов FISH показал, что геном стерляди представляет сложную мозаичную структуру, содержащую диплоидные и тетраплоидные сегменты [22]. При сопоставлении молекулярно-цитогенетических и биоинформатических данных было получено хорошее или приемлемое соответствие результатов для первых двадцати пар хромосом. Для четырех пар хромосом было выявлено существенное несоответствие между физическими размерами хромосом (ARUT17, 21, 22 и 25) и размерами соответствующих им скаффолдов (24, 15, 29 и 18), объясняемое, скорее всего, неполной сборкой. Для большого по размеру скаффолда 11 не удалось установить соответствия с какой-либо хромосомной парой. По-видимому, это артефактный результат.

При сравнении геномов красноухой пресноводной черепахи (*Trachemys scripta elegans*, TSE) и логгерхеда (*Caretta caretta*, CCAR) инструментом D-GENIES были выявлены межхромосомные перестройки, затрагивающие хромосомы TSE2=CCAR2/23/2, TSE4=CCAR6/20 и TSE16=CCAR26/27, также была показана обратная ориентация скаффолдов для гомологичных хромосом TSE1 и CCAR1, TSE4 и CCAR6, TSE9 и CCAR9 и др. Молекулярно-цитогенетическая локализация наборов хромосомоспецифичных проб TSE и коробчатой черепахи (*Terrapene carolina*, TCA) на хромосомах TSE и CCAR хотя и подтвердила наличие трех межхромосомных перестроек, различающих кариотипы TSE и CCAR, но указала на то, что сборка хромосомного уровня не привязана к цитогенетической карте: так, CCAR6 и TSE16 в сборках соответствуют физически CCAR5 и TSE12. Использование микродиссекционных районоспецифичных проб для сравнительного хромосомного пэйнтинга выявило идентичный порядок расположения гомологичных сегментов, например, в гомологичных хромосомах TSE1 и CCAR1, указав на неверную ориентацию скаффолдов в хромосомных сборках.

Сравнительный хромосомный пэйнтинг с пробами домашней собаки продемонстрировал одинаковый порядок гомологичных фрагментов для всех аутосом ластоногих. Высокий уровень консерватизма аутосом был подтвержден при сборке генома байкальской нерпы (*Pusa sibirica,* PSIB) методом Hi-C и сравнении его с геномами пестрой нерпы (*Phoca larga,* PLAR) и обыкновенного

тюленя (*Phoca vitulina,* PVIT) [20]. Было выявлено множество коротких инверсий, и только три инверсии размером около 1 м.п.н. на PLAR6 и PVIT6. Размер выявленных внутрихромосомных перестроек ниже уровня разрешающей способности метода FISH, что подчеркивает важность проведения сравнительного биоинформатического анализа геномов. Однако полученные данные по сборке генома байкальской нерпы до уровня хромосом не содержат сведений о положении центромер, что не позволяет установить, в каком именно плече метацентрической хромосомы находятся выявленные перестройки.
Выводы: Интеграция цитогенетических и биоинформатических данных для корректных сборок геномов хромосомного уровня представителей разных таксонов позвоночных имеет огромное значение, позволяя корректировать ориентацию скаффолдов и устанавливать соответствие с хромосомами кариотипов, описанных молекулярно-цитогенетическими методами.

Relevance of integration of cytogenetics data and chromosome-length genome assemblies at the example of various vertebrates' groups

Romanenko S.[1]*, Beklemisheva V.[1], Perelman P.[1], Andreyushkova D.[1], Marchenko S.[2], Lemskaya N.[1], Biltueva L.[1], Kliver S.[3], Trifonov V.[1], Graphodatsky A.[1]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Independent researcher, Copenhagen, Denmark*

* *rosa@mcb.nsc.ru*

Key words: banding; centromere; chromosome rearrangements; fluorescent *in situ* hybridization; heterochromatin; karyotype; scaffold

Motivation and Aim: Chromosome-length genome assemblies are the gold standard for *de novo* whole genome sequencing [1–3]. The availability of high-quality vertebrate reference genomes led to significant advances in functional, comparative and population genomics [4, 5]. Problematic steps in constructing a reliable reference genome are the correct orientation of contigs and scaffolds relative to each other and the assembly of superscaffolds into structured chromosomal scaffolds, as well as the sequencing and positioning of genomic regions with repeated sequences or segmental duplications [6–8]. Many of the above problems are overcome by integrating genome sequencing data with cytogenetics data. A combined approach that assembles the genome down to the chromosome level is much more successful in answering fundamental biological questions than genomic or cytogenetic approaches alone [9–11]. Unfortunately, integration of genome sequencing data with cytogenetics data is often not carried out, which leads to inconsistency in chromosome numbering even for species for which standardized karyotypes are available, incorrect orientation of scaffolds, incorrect assessment of chromosome sizes due to failure to take into account heterochromatic blocks, and other problems. The purpose of this work is to draw attention to the

importance of understanding and applying a combined approach when performing chromosome-length genome assemblies.

Methods and Algorithms: The methods of classical, differential and molecular cytogenetics were used [12–15]. Microdissection was carried out to obtain chromosome- and region-specific libraries [16]. Preparation of sequencing libraries was performed as previously described [17]. Analysis of sequencing data was carried out using various approaches [17–20].

Results: By pairwise alignment, we identified and visualized syntenic blocks shared by two rodent species: the guinea pig (*Cavia porcellus*), the naked mole rat (*Heterocephalus glaber*), and humans. The high quality of the human genome assembly made it possible to determine the position of centromeres when comparing syntenies. We grouped small syntenies into blocks of at least 5 Mbp to align comparative cytogenetics data and genome assemblies obtained by analyzing sequencing data from Hi-C libraries, which made it possible to link genomic scaffolds to individual chromosomes and orient them relative to centromeres and heterochromatic blocks. Hi-C data made it possible to close all gaps in the comparative maps and identify additional rearrangements that distinguish the karyotypes of the three species. As a result, we integrated bioinformatics and cytogenetic data and adjusted previous comparative maps and genome assemblies of the guinea pig, naked mole rat, and human [19].

In the sterlet, a Eurasian sturgeon fish, (*Acipenser ruthenus*, ARUT) karyotype of 120 chromosomes, the first 14–17 pairs and pairs of small chromosomes carrying NOR genes are easily identified cytogenetically [21]. Using the microdissection, we obtained libraries of individual sterlet chromosomes. These libraries were labeled and used as probes for FISH, and sequenced and aligned to the chromosomal level sterlet genome assembly (GCF_902713425.1/) to link cytogenetic and bioinformatics data. Analysis of FISH results showed that the sterlet genome represents a complex mosaic structure containing diploid and tetraploid segments [22]. When comparing molecular cytogenetic and bioinformatics data, good or acceptable agreement of results was obtained for chromosomes ARUT1-10, 13, 14, as well as the next twenty pairs of chromosomes. For four pairs of chromosomes, a significant discrepancy was detected between the physical sizes of the chromosomes (ARUT17, 21, 22 and 25) and the sizes of their corresponding scaffolds (24, 15, 29 and 18), most likely explained by incomplete assembly. For the large scaffold 11, it was not possible to establish a correspondence with any chromosomal pair.

When comparing the genomes of two turtles: the red-eared slider (*Trachemys scripta elegans*, TSE) and the loggerhead sea turtle (*Caretta caretta*, CCAR) using the D-GENIES tool, interchromosomal rearrangements TSE2=CCAR2/23/2, TSE4=CCAR6/20, and TSE16=CCAR26/ 27 were identified, the reverse orientation of scaffolds was also shown for homologous chromosomes TSE1 and CCAR1, TSE4 and CCAR6, TSE9 and CCAR9, etc. Molecular cytogenetic localization of sets of chromosome-specific probes TSE and the eastern box turtle (*Terrapene carolina*, TCA) on the chromosomes of TSE and CCAR, although confirmed the presence of the three interchromosomal rearrangements distinguishing the TSE and CCAR karyotypes, but indicated that the chromosome-length genome assembly is not tied to the cytogenetic map: thus, CCAR6 and TSE16 in the assembly physically correspond to CCAR5 and TSE12. The use of microdissection region-specific probes for comparative chromosomal painting revealed an identical order of arrangement of homologous segments, for

example, in the homologous chromosomes of TSE1 and CCAR1, indicating the incorrect orientation of scaffolds in the turtles' chromosomal assemblies.

Comparative chromosomal painting with the domestic dog probes demonstrated the same order of homologous fragments for all pinniped autosomes (Carnivora, Mammalia). A high level of autosomal conservatism was confirmed by assembling the genome of the Baikal seal (*Pusa sibirica*, PSIB) using the Hi-C method and comparing it with the genomes of the pied seal (Phoca larga, PLAR) and the common seal (*Phoca vitulina*, PVIT) [20]. Many short inversions and three inversions approximately 1 Mb in size on PLAR6 and PVIT6 were identified. The size of the identified intrachromosomal rearrangements is below the resolution level of FISH, which emphasizes the importance of conducting comparative bioinformatics analysis of genomes. However, the data obtained on the assembly of the Baikal seal genome to the chromosome level do not contain information about the position of the centromeres, which does not allow us to establish in which arm of the metacentric chromosome the identified rearrangements were located.

Conclusion: Integration of cytogenetic and bioinformatics data for correct chromosome-length genome assemblies is of great importance, making it possible to correct the orientation of scaffolds and to match chromosome scaffolds with chromosomes described by molecular cytogenetic methods in representatives of different vertebrate taxa.

Список литературы/References

1. Rhie A. et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature.* 2021;592(7856):737-746. doi 10.1038/s41586-021-03451-0
2. Kim J. et al. Reference-assisted chromosome assembly. *Proc Natl Acad Sci USA.* 2013;110(5):1785-1790. doi 10.1073/pnas.1220349110
3. A reference standard for genome biology. *Nat Biotechnol.* 2018;36(12):1121. doi 10.1038/nbt.4318
4. Garner B.A. et al. Genomics in Conservation: Case Studies and Bridging the Gap between Data and Application. *Trends Ecol Evol.* 2016;31(2):81-83. doi 10.1016/j.tree.2015.10.009
5. Supple M.A., Shapiro B. Conservation of biodiversity in the genomics era. *Genome Biol.* 2018;19:13. doi 10.1186/s13059-018-1520-3
6. Zimin A.V. et al. Mis-assembled "segmental duplications" in two versions of the *Bos taurus* genome. *PLoS One*. 2012;7(8):e42680. doi 10.1371/journal.pone.0042680
7. Hills M. et al. Construction of whole genomes from scaffolds using single cell strand-seq data. *Int J Mol Sci.* 2021;22(7):3617. doi 10.3390/ijms22073617
8. Luo J. et al. A comprehensive review of scaffolding methods in genome assembly. *Brief Bioinform.* 2021;22(5):bbab033. doi 10.1093/bib/bbab033
9. Deakin J.E. et al. Chromosomics: Bridging the Gap between Genomes and Chromosomes. *Genes (Basel).* 2019;10(8):627. doi 10.3390/genes10080627
10. Lewin H.A. et al. Precision nomenclature for the new genomics. *GigaScience.* 2019;8(8):giz086. doi 10.1093/gigascience/giz086
11. Iannucci A. et al. Bridging the gap between vertebrate cytogenetics and genomics with Single-Chromosome Sequencing (ChromSeq). *Genes (Basel).* 2021;12(1):124. doi 10.3390/genes12010124
12. Lemskaya N.A. et al. A combined banding method that allows the reliable identification of chromosomes as well as differentiation of AT- and GC-rich heterochromatin. *Chromosome Res.* 2018;26(4):307-315. doi 10.1007/s10577-018-9589-9
13. Seabright M. A rapid banding technique for human chromosomes. *Lancet.* 1971;2(7731):971-972. doi 10.1016/s0140-6736(71)90287-x
14. Sumner A.T. A simple technique for demonstrating centromeric heterochromatin. *Exp Cell Res.* 1972;75(1):304-306. doi 10.1016/0014-4827(72)90558-7
15. Yang F., Graphodatsky A.S. Animal probes and ZOO-FISH. In: Liehr T. (Ed.). Fluorescence In Situ Hybridization (FISH). Springer Protocols Handbooks, 2017. doi 10.1007/978-3-662-52959-1_42
16. Romanenko S., Trifonov V. Generation of microdissection-derived painting probes from single copy chromosomes. In: Liehr T. (Ed.). Cytogenetics and Molecular Cytogenetics. CRC Press, 2022. doi 10.1201/9781003223658

17. Andreyushkova D.A. et al. Next generation sequencing of chromosome-specific libraries sheds light on genome evolution in paleotetraploid sterlet (*Acipenser ruthenus*). *Genes (Basel).* 2017;8(11):318. doi 10.3390/genes8110318
18. Makunin A.I. et al. Contrasting origin of B chromosomes in two cervids (Siberian roe deer and grey brocket deer) unravelled by chromosome-specific DNA sequencing. *BMC Genomics.* 2016;17(1):618. doi 10.1186/s12864-016-2933-6
19. Romanenko S.A. et al. Integration of fluorescence *in situ* hybridization and chromosome-length genome assemblies revealed synteny map for guinea pig, naked mole-rat, and human. *Sci Rep.* 2023;13(1):21055. doi 10.1038/s41598-023-46595-x
20. Yakupova A. et al. Chromosome-length assembly of the Baikal seal (*Pusa sibirica*) genome reveals a historically large population prior to isolation in Lake Baikal. *Genes.* 2023;14(3):619. doi 10.3390/genes14030619
21. Romanenko S.A. et al. Segmental paleotetraploidy revealed in sterlet (*Acipenser ruthenus*) genome by chromosome painting. *Mol Cytogenet.* 2015;8:90. doi 10.1186/s13039-015-0194-8
22. Trifonov V.A. et al. Evolutionary plasticity of acipenseriform genomes. *Chromosoma.* 2016;125:661-668. doi 10.1007/s00412-016-0609-2

doi 10.18699/bgrs2024-01-08

Структурная организация эукариотического генома на примере *C. merolae*

Руденко В.*, Коротков Е.
ФИЦ Биотехнологии РАН, Москва, Россия
** v.m.rudenko@gmail.com*

Ключевые слова: дисперсные повторы; повторы в ДНК; семейства повторов; позиционно-весовая матрица; *Cyanidioschyzon merolae*

Мотивация и цель: Как известно, эукариотическая ДНК информационно избыточна. Она содержит большое число повторов, тандемных и дисперсных. В данной работе ставилась задача выявления и последующего изучения дисперсных повторов в геноме эукариотического организма *Cyanidioschyzon merolae*. Для обнаружения повторов использовался IP-метод [1], который ранее успешно применялся для поиска повторов в геномах бактерий. Обсуждаются достоинства и недостатки IP-метода для поиска дисперсных повторов, а также возможное функциональное значение повторов.
Методы и алгоритмы: Поиск дисперсных повторов проводился в геноме примитивного растительного организма – красной водоросли *Cyanidioschyzon merolae* (https://plants.ensembl.org/Cyanidioschyzon_merolae/Info/Annotation/#assembly).
Геном содержит 20 хромосом и имеет длину 16,546,747 н.п. Также исследовались последовательности ДНК хлоропласта и митохондрии. Геном *C. merolae* содержит относительно мало генов – 5,331, из которых аннотировано 4,984; практически все гены лишены интронов.
Метод IP, используемый в данной работе для поиска дисперсных повторов, имеет несколько преимуществ. Во-первых, он относится к категории *de novo* методов, т. е. не использует никакой информации о строении повтора. Во-вторых, он способен выявлять сильноразмытые дисперсные повторы. Это означает, что число замен на одно основание x между последовательностями отдельных представителей семейства повторов может достигать 1.5. В то же время, как было показано в [1], мощность традиционных методов аннотации повторов, таких как RepeatMasker [2], Repeat Detector (RED) [3], RECON [4] и т. д., ограничивается $x<=1.0$. Такая чувствительность IP-метода достигается за счет того, что для поиска подобий используется обобщенный профиль семейства повторов, который настраивается на конкретный геном в ходе итерационной процедуры. Также учитываются парные корреляции соседних оснований в профиле семейства повторов. В-третьих, метод строит консенсус семейства повтора по множественному выравниванию последовательностей, входящих в него. Это позволяет охарактеризовать семейство и сравнить его с существующей таксономией повторов.
Результаты: В геноме *C. merolae* было выделено 20 семейств повторов, общее число повторов составляет 33,938 [5]. Повторы равномерно распределены по хромосомам: около 2 повторов приходятся на каждые 1000 н.п., они покрывают 72.64 % генома. Длины повторов варьируют от 108 до 600 н.п., средняя длина повтора – 523 н.п.

Определялось пересечение повторов с генами. Оказалось, что 86 % генов пересекаются с повторами таким образом, что длина пересечения превышает половину длины гена. Поскольку гены в ядерной ДНК занимают около 44 % хромосомы, то можно сделать вывод, что найденные нами повторы предпочитают располагаться в кодирующих районах ДНК.

Полученные данные по дисперсным повторам мы сравнили с имеющейся по геному *C. merolae* аннотацией с сайта http://plants.ensembl.org. Для аннотации повторов там применялись программные средства Dust [6], TRF [7], RepeatMasker [2] и RED [3]. Повторы, идентифицируемые программами Dust и TRF, – это low complexity regions и тандемные повторы; они не являлись целью нашего исследования. Поэтому для корректного сравнения были выбраны только дисперсные повторы, определяемые RepeatMasker и RED. RepeatMasker относится к категории методов, которые используют существующие базы данных по дисперсным повторам. Для аннотации *C. merolae* применялись базы данных Redat и Repbase. RepeatMasker и RED детектировали 20,320 повторов, которые покрывают 28.09 % генома. Длина повторов варьируется в широких пределах: от 15 до 19,220 н.п. Однако в целом это более короткие повторы, чем найденные методом IP, средняя длина их составляет всего 260 н.п.

Также было изучено пересечение классов аннотированных повторов с найденными нами семействами. Пусть $v(i,j)$ – число пересечений i аннотированного класса и j семейства повторов найденного IP-методом. Определим величину $v`(i,j)$:

$$v'(i,j) = \frac{\left(v(i,j) - np(i,j)\right)}{\sqrt{np(i,j)\left(1 - p(i,j)\right)}}\ , \qquad (1)$$

где $x(i) = \sum_j v(i,j),\ y(j) = \sum_i v(i,j)$, $p(i,j) = x(i)y(j)/n,\ n = \sum_i \sum_j v(i,j)$

$v`(i,j)$ имеет приближенно стандартное нормальное распределение и показывает степень корреляции между существующими классами и нашими семействами. Большие позитивные значения матрицы $v`(i,j)$ были обнаружены для семейств № 11–13 и классов LTR-повторов Copia и Gypsy.

Для всех 20 семейств повторов на основании множественного выравнивания последовательностей, по которым генерировался профиль семейства, строились консенсусы семейств. Визуальные представления консенсусов были получены с помощью программы Weblogo [8]. Консенсус для самого крупного 1-го семейства повторов изображен на рис. 1. Позиции в консенсусе представлены нуклеотидами, высота которых пропорциональна частоте их встречаемости, преобразованной в биты. Можно видеть, что данное семейство достаточно размыто; некоторые участки внутри повторов жестко определены, другие же относительно случайны.

Выводы: Применение IP-метода к анализу генома *C. merolae* показало, что в этом геноме можно выделить 20 семейств повторов. Семейства включают в себя около 34 тысяч дисперсных повторов, которые покрывают более 72 % генома. Это гораздо больше, нежели было определено ранее. Однако этот факт не умаляет достоинств других программных средств аннотации повторов, таких как RepeatMasker и RED. Они ориентированы на поиск коротких сильногомологичных повторов, которые наш метод определяет не в полной мере. Тем не менее некоторые классы аннотированных повторов, как было показано, коррелируют с найденными IP-методом семействами. Мы рекомендуем для

полноценной геномной аннотации повторов использовать совокупность различных биоинформационных подходов.

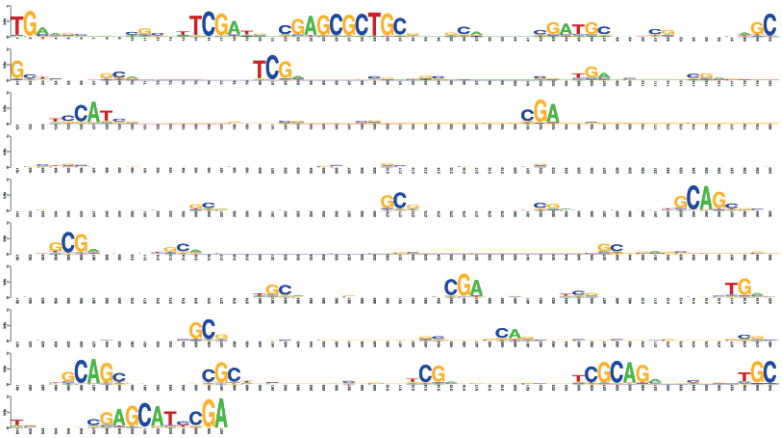


Рис. 1. Консенсус 1 семейства повторов

Выявленные в работе повторы покрывают значительную часть генома. Более правильно было бы назвать их структурными элементами или некоторой разметкой, которая накладывается на весь ядерный геном. В дальнейшем было бы интересно понять, является ли такая разметка уникальной в пределах биологического вида, или же это общее свойство генома.

Мы предполагаем, что подробная структурная организация может быть связана с компактизацией хроматина. Возможно, консервативные участки в повторах представляют собой сайты связывания гистонов в процессе образования нуклеосом.

Финансирование: Исследование поддержано грантом РНФ (№ 24-24-00031).

Structural organization of the eukaryotic genome by example of *C. merolae*

Rudenko V.*, Korotkov E.

Research Center of Biotechnology of RAS, Moscow, Russia

** v.m.rudenko@gmail.com*

Key words: dispersed repeats; repeated DNA; position weight matrix; repeat family; *Cyanidioschyzon merolae*

Motivation and Aim: As is known, eukaryotic DNA is informationally redundant. It contains a large number of tandem and dispersed repeats. The aim of this work is identifying and studying of dispersed repeats in the genome of the eukaryotic organism *Cyanidioschyzon merolae*. To detect repeats, we used the IP method [1], which was previously successfully used to search for repeats in bacterial genomes. The advantages and disadvantages of the IP method for searching for dispersed repeats are discussed, as well as the possible functional significance of the repeats.

Methods and Algorithms: The search for dispersed repeats was carried out in the genome of a primitive plant organism the red alga *C. merolae* (https://plants.ensembl.org/Cyanidioschyzon_merolae/Info/Annotation/#assembly). The genome contains 20 chromosomes; the length is 16,546,747 bp. Chloroplast and mitochondrial sequences were considered too. The *C. merolae* genome has only 5,331 genes 4,984 of which are annotated. Almost all genes lack introns.

The IP method used to search for dispersed repeats has several advantages. Firstly, it belongs to the category of *de novo* methods, so it does not use any information about the structure of the repeats. Secondly, it is capable of identifying highly dispersed repeats; the degree of similarity x between different repeats from the family can reach 1.5. It was shown in [1] the power of traditional repeats annotation methods such as RepeatMasker [2] and Repeat Detector (RED) [3], RECON [4], etc. is limited by $x<=1.0$. This sensitivity of the IP method is achieved due to the fact that a generalized repeat family profile is used to search for similarities, which is tuned to a specific genome during an iterative procedure. Pairwise correlations of neighboring bases in the repeat family profile are also taken into account. Thirdly, the method builds a consensus of the repeat family based on multiple alignment of the sequences included in it. This allows to characterize the family and compare it to existing repeat taxonomy.

Results: 20 repeat families have been identified in the *C. merolae* genome, with a total number of repeats of 33,938 [5]. The repeats are uniform distributed along the chromosomes, about 2 repeats for every 1000 bp and cover 72.64 % of the genome. Repeat lengths vary from 108 to 600 bp the average repeat length is 523 bp.

The intersection of repeats with genes was determined. It turned out that 86 % of genes intersect with repeats, such that the length of the intersection exceeds half the length of the gene. Since genes in nuclear DNA occupy about 44 % of the chromosome, we can conclude that the repeats prefer to be located in the coding regions of the DNA.

We compared the obtained data on dispersed repeats with the annotation available for the *C. merolae* genome from the website http://plants.ensembl.org. At this site repeats were annotated by the software tools Dust [6], TRF [7], RepeatMasker [2] and RED [3]. The repeats identified by the Dust and TRF programs are low complexity regions and tandem repeats; they were not the purpose of our study. Therefore, only dispersed repeats identified by RepeatMasker and RED were selected for comparison. RepeatMasker belongs to the category of methods that use existing databases of dispersed repeats. The Redat and Repbase databases were used to annotate *C. merolae*. RepeatMasker and RED detected 20,320 repeats, which covered 28.09 % of the genome. The length of the repeats varies widely: from 15 to 19,220 bp. However, in general, these are shorter repeats than those found by the IP method; their average length is only 260 bp.

We also examined the intersection of classes of annotated repeats with the families we found. Let $v(i,j)$ be the number of intersections of the i annotated class and the j family of repeats found by the IP method. Let's define the value $v`(i,j)$:

$$v'(i,j) = \frac{\left(v(i,j) - np(i,j)\right)}{\sqrt{np(i,j)\left(1 - p(i,j)\right)}}, \qquad (1)$$

where $x(i) = \sum_j v(i,j),\ y(j) = \sum_i v(i,j)$, $p(i,j) = x(i)y(j)/n,\ n = \sum_i \sum_j v(i,j)$

$v`(i,j)$ has an approximately standard normal distribution and shows the degree of correlation between existing classes and our families. Large positive values of the matrix

v`(i,j), indicating the presence of a correlation, were found between families No. 11–13 and the LTR repeat classes Copia and Gypsy.

For all 20 repeat families consensuses were constructed in text and numerical form Consensus based on multiple sequence alignment from which the family profile was generated. Visual representations of consensuses were obtained using the Weblogo program [8]. The consensus for the largest 1st repeat family is shown in Figure 1. Consensus positions are represented by nucleotides whose height is proportional to their frequency of occurrence converted to bits. It can be seen that this family is quite diverged, there is no strong homology between individual repeats over their full length. Some regions within repeats are strictly defined, while others are relatively random.

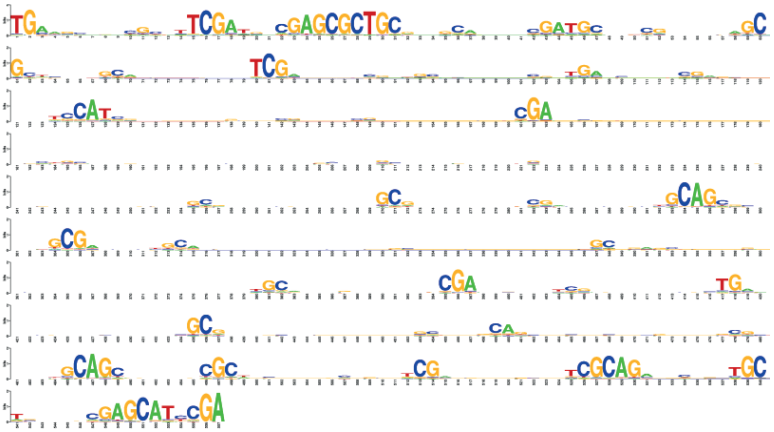


Fig. 1. Consensus for the first repeat family

Conclusion: The application of the IP method to the analysis of the *C. merolae* genome showed that 20 repeat families can be identified. The families include about 34 thousand dispersed repeats, which cover more than 72 % of the genome. This is much more than was previously determined. However, this fact does not detract from the merits of other repeat annotation software tools, such as RepeatMasker and RED. They are focused on searching for short, strongly homologous repeats, which our method does not fully detect. However, some classes of annotated repeats are correlating with families found by the IP method. We recommend using a combination of different bioinformatics approaches for complete genomic annotation of repeats.

The repeats identified in the work cover a significant part of the genome. It would be more correct to call them structural elements or some kind of marking that is superimposed on the entire nuclear genome, including both coding and non-coding regions. In the future, it would be interesting to understand whether such marking is unique within a biological species, or whether this is a general property of the genome.

We hypothesize that the structural organization of the DNA may be related to chromatin compaction. It is possible that conserved regions in the repeats represent histone binding sites in the formation of nucleosomes process. By changing the location of nucleosomes, the DNA molecule gets a chance to switch to the translation of alternative genes to adapt to changes in external conditions or ensure tissue specificity.

Funding: The study is supported by RSF (No. 24-24-00031).

Список литературы/References

1. Korotkov E., Suvorova Y., Kostenko D., Korotkova M. Search for dispersed repeats in bacterial genomes using an iterative procedure. *Int J Mol Sci*. 2023;24:10964. doi 10.3390/IJMS241310964/S1
2. RepeatMasker Home page. Available online: https://repeatmasker.org/ (accessed on Aug 11, 2022)
3. Girgis H.Z. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinformatics*. 2015;16:227. doi 10.1186/S12859-015-0654-5/TABLES/4
4. Bao Z., Eddy S.R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res*. 2002;12:1269-1276. doi 10.1101/GR.88502
5. Rudenko V., Korotkov E. Study of Dispersed Repeats in the Cyanidioschyzon merolae Genome. *Int J Mol Sci*. 2024;25:4441. doi 10.3390/IJMS25084441
6. Morgulis A., Gertz E.M., Schäffer A.A., Agarwala R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J Comput Biol*. 2006;13:1028-1040. doi 10.1089/CMB.2006.13.1028
7. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573-580. doi 10.1093/nar/27.2.573 1999;27
8. Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. WebLogo: a sequence logo generator. *Genome Res*. 2004;14:1188-1190. doi 10.1101/GR.849004

doi 10.18699/bgrs2024-1.1-09

Локальная структура ДНК базальных промоторов генов тРНК эукариот

Савина Е.А.[1, 2]*, Орлов Ю.Л.[1, 3], Лебедев Г.С.[1], Анашкина А.А.[1, 2], Ильичёва И.А.[2]

[1] *ФГАОУ ВО Первый Московский государственный медицинский университет им. И.М. Сеченова Министерства здравоохранения Российской Федерации (Сеченовский Университет), Москва, Россия*

[2] *Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

[3] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

** savina_e_a_1@staff.sechenov.ru*

Ключевые слова: транскрипция тРНК; TATA-связывающий белок; ультразвуковое расщепление ДНК; расщепление ДНКазой I; локальная структура ДНК

Мотивация и цель: Три ДНК-зависимые РНК-полимеразы осуществляют транскрипцию всех типов молекул РНК в клетках эукариот. Это РНК-полимераза I (Pol I), которая транскрибирует предшественник рРНК, РНК-полимераза II (Pol II), которая транскрибирует предшественники информационных РНК и различных регуляторных некодирующих РНК, и РНК-полимераза III (Pol III), которая транскрибирует предшественники малых РНК, включая все тРНК, 5SpРНК, 7SLРНК, малых ядерных РНК U6 [1], а также мобильные генетические элементы, называемые SINE [2]. В растениях дополнительно присутствуют еще две РНК-полимеразы [3]. Ключевым фактором транскрипции Pol II является белок (TBP), распознающий TATA-бокс в базальных промоторах генов матричной РНК. Ранее мы показали, что позиция связывания его в окрестностях –28 п. н. от старта транскрипции определяется локальными структурными свойствами ДНК на этом участке [4–6]. Этот белок присутствует и в транскрипционных комплексах Pol I и Pol III. Целью данной работы было выявление аналогичного сайта связывания TBP в области, предшествующей генам тРНК в геномах эукариот.

Методы и алгоритмы: Мы использовали коллекцию тРНК 7 эукариот: *S. pombe 972h-* (171 ген), *S. cerevisiae S288c* (273 гена), *A. thaliana TAIR 10* (642 гена), *D. melanogaster* (*BDGP Rel. 6/dm6*) (290 генов). *C. elegans W BceI235/ce 11* (580 генов), *Mus musculus GRCm39/mm39* (400 генов) и *H. sapiens GRCh38/hg38* (429 генов), находящуюся в базе данных геномных тРНК (GtRNAdb). Эта база содержит гены тРНК, выявленные на полных геномах с помощью компьютерного поиска программой tRNAscan-SE. Мы проанализировали нуклеотидный текст и изменения локальной пространственной структуры фрагментов длиной 80 н. п. в окрестностях начала генов тРНК (–50/+30). Были определены профили частот встречаемости мононуклеотидов, их лого-представления, а также профили изменений физических и структурных параметров интенсивностей ультразвукового расщепления, и расщепления ДНКазой I. Для этой цели нами была написана программа на Python 3.10.

Результаты: Оказалось, что профили, характеризующие тексты участков, предшествующих генам (от –50 до –1), существенно различаются у разных представителей эукариот, а начальные участки самих генов тРНК (от +1 до +30) у всех эукариот имеют сходные последовательности и одинаковый консенсус бокса A.

В отличие от этого профили структурных и физических параметров, включая характеристики интенсивностей расщеплений ДНК ультразвуком и с помощью фермента ДНКазы I, выявляют сходство локальной структуры участков, предшествующих генам (-50/-1). Эти структурные особенности, которые наиболее выражены у *S. pombe 972h-*, показаны на рис. 1. Сравнение структурных профилей всех семи изученных организмов указывают на расширение малой бороздки ДНК и низкую энергию изгиба в сторону несахарного желоба, немного варьирующую на всем протяжении, вплоть до начала гена. Такие особенности структуры и динамики ДНК вблизи генов тРНК существенно отличают их от базальных промоторов матричной РНК.

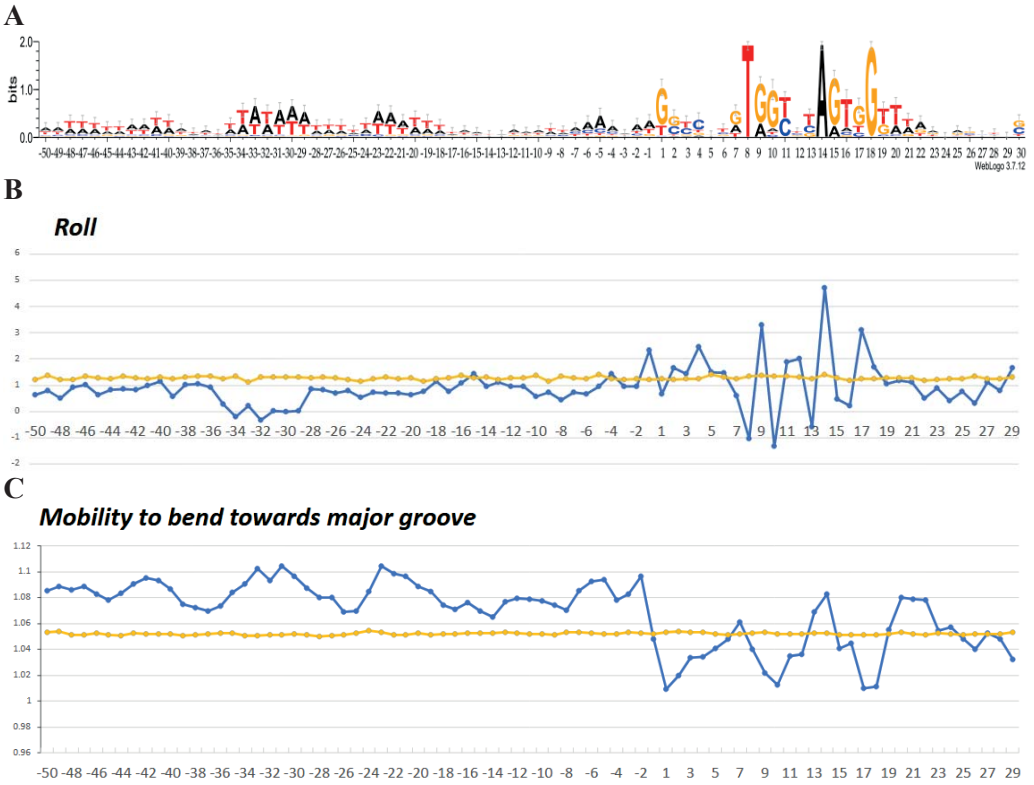


Рис. 1. A. Лого-представление областей от –50 до +30 для 171 гена тРНК *S. pombe 972h-*. B. Структурные параметры Roll. C. Mobility to bend towards major groove для областей от –50 до +30 генов тРНК *S. pombe 972h-* (синий цвет) и случайных последовательностей (желтый)

Выводы: Особенности структуры ДНК на участках базальных промоторов генов тРНК, которые заметно отличают их от промоторных участков генов матричной РНК, транскрибируемых РНК-полимеразой II, позволяют предположить, что при транскрипции генов тРНК TATA-связывающий белок не играет ключевой роли в формировании транскрипционного комплекса РНК-полимеразы III.

Финансирование: Исследование поддержано грантом Российского научного фонда № 24-24-00563 на тему «Разработка цифровых образовательных программ в биомедицине».

Local structure of basal DNA promoters in eukaryotic tRNA genes

Savina E.[1, 2]*, Orlov U.[1, 3] , Lebedev G.[1], Anashkina A.[1, 2], Il'icheva I.[2]

[1] *Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

[2] *V.A. Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

** savina_e_a_1@staff.sechenov.ru*

Key words: tRNA transcription; TBP; ultrasonic cleavage; DNase I cleavage; local structure of DNA

Motivation and Aim: Transcription of all types RNA molecules is carried out by three DNA-dependent RNA polymerases in eukaryotic cells. RNA polymerase I (Pol I) synthesizes the rRNA precursor, RNA polymerase II (Pol II) synthesizes messenger RNAs and various regulatory non-coding RNAs, and RNA polymerase III (Pol III) synthesizes small RNAs, including all tRNAs, 5SrRNAs, 7SLRNAs, U6 small nuclear RNAs [1], as well as mobile genetic elements referred to as SINEs [2]. In plants also function two additional RNA polymerases [3]. A key general transcription factor in Pol II machinery is TATA-binding protein (TBP) that recognizes the TATA box in the basal promoters of RNA genes. We have previously shown that its binding position in the vicinity of –28 bp from the transcription start (TSS) is determined by local structural properties of DNA at this site [4–6]. TBP is the component of Pol I and Pol II transcription complexes also. The aim of this work was to find analogical binding site of TBP in the region upstream tRNA genes in eukaryotic genomes.

Methods and Algorithms: We utilized the tRNA collection of 7 eukaryotic organisms: *S. pombe 972h-* (171 genes), *S. cerevisiae S288c* (273 genes), *A. thaliana TAIR 10* (642 genes), *D. melanogaster* (*BDGP Rel. 6/dm6*) (290 genes). *C. elegans W BceI235/ce 11* (580 genes), *Mus musculus GRCm39/mm39* (400 genes) and *H. sapiens GRCh38/hg38* (429 genes), from genomic tRNA database (GtRNAdb). It contains tRNA genes identified on full genomes by tRNAscan-SE – computer search program. We analyzed the nucleotide text and changes in the local spatial structure of fragments 80 np in length in the vicinity of tRNA gene starts (–50/+30). Profiles of mononucleotide occurrence frequencies, their logo-presentation, and profiles of changes in physical and structural parameters, ultrasonic cleavage and DNAase I cleavage intensities were determined. For this purpose, we wrote a program in Python 3.10.

Results: It turned out that the profiles characterizing the texts of the sites preceding genes (from –50 to –1) differ significantly among the representatives of eukaryotes, while the sequences of tRNA genes themselves (from +1 to +30) in all eukaryotes are similar, as well as their consensus of box A. In contrast, all profiles of structural and physical parameters, including characterization of DNA cleavage intensities by ultrasound and by the DNAase I enzyme, reveal similarities in the local structure of the regions preceding the genes (–50/–1). These structural features, which are most pronounced in *S. pombe 972h-*, are shown in Figure 1. Comparison of the structural profiles of all seven organisms studied indicate a widening of the minor groove of DNA and a low bending energy towards the major groove, slightly varying up to the gene start. These features of

DNA structure and dynamics near tRNA genes essentially distinguish them from the basal promoters of matrix RNA.

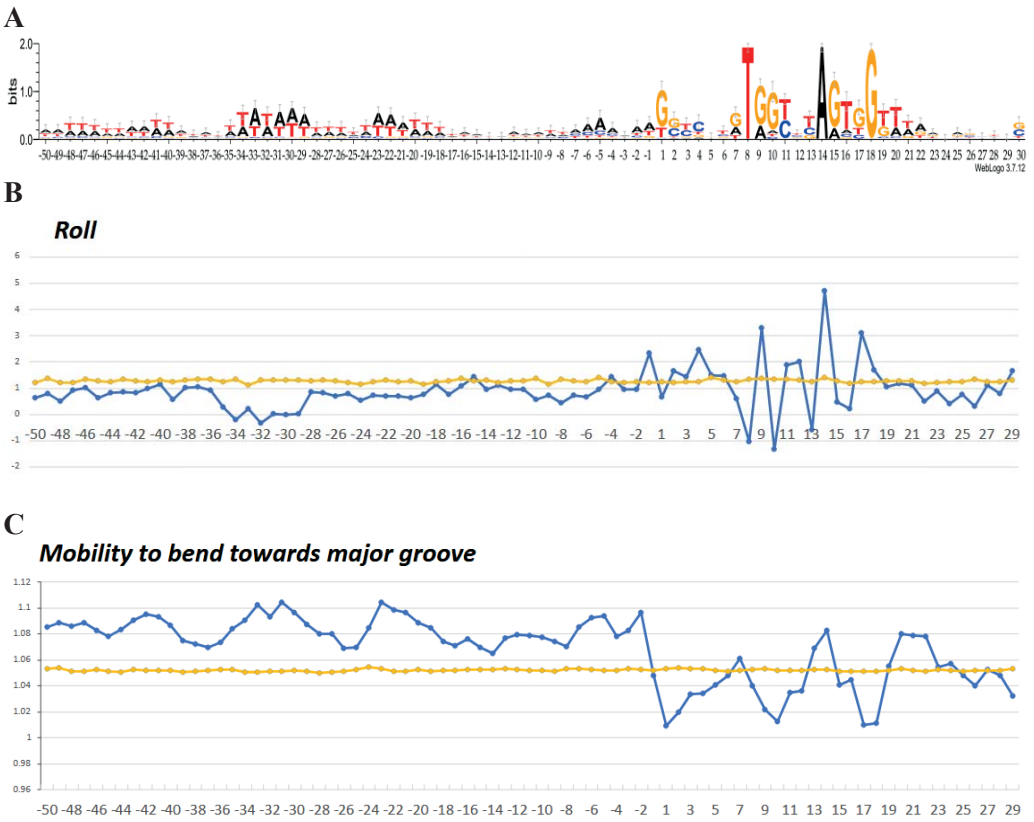


Fig. 1. A. Logo representation of the –50 to +30 regions for 171 *S. pombe 972h-* tRNA genes. B. Structural parameters Roll. C. Mobility to bend toward major groove for the –50 to +30 regions of *S. pombe 972h-* tRNA genes (blue) and random sequences (yellow)

Conclusion: These features of local structure of DNA in the vicinity of tRNA genes, which markedly distinguish them from the promoter sites of mRNA genes transcribed by RNA polymerase II, suggest that TATA-binding protein doesn't play a key role in the transcriptional complex formation of RNA polymerase III.

Funding: The study is supported by Russian Science Foundation (No. 24-24-00563) on the topic "Development of digital educational programs in biomedicine".

Список литературы/References

1. Girbig M., Misiaszek A.D., Müller C.W. Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nat Rev Mol Cell Biol.* 2022;23(9):603-622
2. Vassetzky N.S., Kramerov D.A. SINEBase: a database and tool for SINE analysis. *Nucleic Acids Res.* 2013;41(D1):D83-D89
3. Zhou M., Law J.A. RNA Pol IV and V in gene silencing: rebel polymerases evolving away from Pol II's rules. *Curr Opin Plant Biol*. 2015;27:154-164
4. Il'icheva I.A., Khodikov M.V., Poptsova M.S., Nechipurenko D.Y., Nechipurenko Y.D., Grokhovsky S.L. Structural features of DNA that determine RNA polymerase II core promoter. *BMC Genomics.* 2016;17:1-21
5. Melikhova A.V., Anashkina A.A., Il'icheva I.A. Evolutionary invariant of the structure of DNA double helix in RNAP II Core Promoters. *Int J Mol Sci.* 2022;23:10873. doi 10.3390/ijms231810873
6. Savina E.A., Shumilina T.G., Tumanyan V.G., Anashkina A.A., Il'icheva I.A. Core promoter regions of antisense and long intergenic non-coding RNAs. *Int J Mol Sci.* 2023;24(9):8199

doi 10.18699/bgrs2024-1.1-10

Связь информационной емкости и таксономии геномов оспы

Сенашова М.[1]*, Садовский М.[1, 2, 3]

[1] *Институт вычислительного моделирования СО РАН, Красноярск, Россия*

[2] *ФСНКЦ ФМБА России, Красноярск, Россия*

[3] *Сибирский федеральный университет, Красноярск, Россия*

* *msen@icm.krasn.ru*

Ключевые слова: нуклеотидная последовательность; частота; условная энтропия; главные компоненты; кластер

Мотивация и цель: Анализ статистических свойств нуклеотидных последовательностей является основной задачей биофизики, биоинформатики и молекулярной биологи. Множество работ было посвящено различным аспектам этого направления исследований (см., например, [1–4]). В данной работе рассмотрена связь между информационной емкостью частотных словарей нуклеотидных последовательностей толщиной от 2 до 20 символов и биологическими свойствами этих последовательностей. Каждой нуклеотидной последовательности ставится в соответствие точка в 18-мерном пространстве. Координатами точки являются условные энтропии, вычисленные для данной последовательности. Проанализирован состав кластеров, образованных проекциями точек из 18-мерного пространства в пространство первых трех главных компонент.

Методы и алгоритмы: Мы будем использовать частотные словари – совокупность всех фрагментов нуклеотидной последовательности фиксированной длины, встречающихся в ней, с указанием их частоты [5–7]. Введем основные понятия. Рассмотрим нуклеотидную последовательность длины N, состоящую из символов *A*, *C*, *G*, *T*. Любую последовательность символов длины q будем называть словом ω. Пусть n_ω – количество копий этого слова в нуклеотидной последовательности; тогда пару $\langle \omega, n_\omega \rangle$ будем называть (конечным) словарем *Wq* толщины *q*. Заменив количество копий частотой $f_\omega = \frac{n_\omega}{N}$, получаем частотный словарь *Wq* той же толщины. В работе [8] была изучена информационная емкость символьных последовательностей как условная энтропия частотных словарей толщины *q*, вычисленная по частотным словарям меньшей толщины. Мы используем этот подход для вычисления информационной емкости генетических последовательностей для разной толщины словаря. Информационная емкость или условная энтропия словаря толщины q может быть вычислена по формуле: $\overline{S} = 2S_{q-1} - S_q - S_{q-2}$ либо $\overline{S} = 2S_1 - S_2$ для $q = 2$, где $S_q = -\sum_{f_\omega} f_\omega \ln f_\omega$ – абсолютная энтропия частотного словаря толщины *q*. В нашем случае $q \le 20$. То есть для каждого генома мы получаем 18 значений условной энтропии.

Результаты: Были рассмотрены 70 геномов, относящихся к семейству Poxviridae, размещенных в GenBank. Для этих геномов были вычислены условные энтропии

для словарей толщиныойот 1 до 20. Набор условных энтропий для каждого генома рассматриваем как точку в 18-мерном пространстве. Далее в программе VidaExpert (http://bioinfo-out.curie.fr/projects/vidaexpert/) проецируем полученное множество точек в пространство первых трех главных компонент. Были проанализированы кластеры, состоящие из точек, чьи значения условной энтропии близки в метрике евклидова пространства. Были обнаружены следующие кластеры: оспа насекомых (AF063866 *Melanoplus sanguinipes entomopoxvirus O isolate Tucson*, AF250284 *Amsacta moorei entomopoxvirus*, AP013055 *Anomala cuprea entomopoxvirus DNA*, HF679131 *Adoxophyes honmai enomopoxvirus L*, HF679132 *Choristoneura biennis entomopoxvirus L*, HF679133 *Choristoneura rosaceana entomopoxvirus L*, HF679134 *Mythimna separata entomopoxvirus L*). Геном KR095315 *Diachasmimorpha longicaudata entomopoxvirus* не попал в этот кластер, так как *GC*-состав этого генома равен 0.3, а *GC*-составы остальных геномов оспы насекомых имеют значения около 0.2. Оспа птиц (AF198100 *Fowlpox virus*, AY318871 *Canarypox virus strain ATCC VR-111*, KJ801920 *Pigeonpox virus isolate FeP2*, KJ859677 *Penguinpox virus isolate PSan92*, MF678796 *Flamingopox virus FGPVKD09*, MK903864 *Magpiepox virus*, MT799800 *Cheloniid poxvirus 1*, MW365933 *Albatrosspox virus strain SAN97-0665NZ*, MW485973 *Magpiepox virus 2 isolate 62-11-06-2000-ANU*). В этот же кластер попал геном оспы черепахи MT799800 *Cheloniid poxvirus 1*. Кластер, в состав которого входят геномы, относящиеся к роду *Orthopoxvirus* (AF482758 *Cowpox virus strain Brighton Red*, AY009089 *Camelpox virus CMS*, DQ437594 *Taterapox virus strain Dahomey 1968*, HM172544 *Monkeypox virus strain Zaire 1979-005*, KP143769 *Raccoonpox virus*, KU749310 *Skunkpox virus strain WA*, KU749311 *Volepox virus strain CA*, L22579 *Variola major virus strain Bangladesh-1975*, MH607141 *Akhmeta virus isolate Akhmeta 2013-88*, MH816996 *Orthopoxvirus Abatino*, MN240300 *Borealpox virus*, OP526861 Monkeypox virus isolate MPXV USA 2022 FL0019, X69198 *Variola virus DNA*, Y16780 *Variola minor virus*). Кластер, в который входят геномы вирусов оспы парнокопытных (AF325528 *Lumpy skin disease virus NI-2490 isolate Neethling 2490*, AF410153 *Swinepox virus isolate 17077-99*, AY077832 *Sheeppox virus 10700-99 strain TU-V02127*, AY077835 *Goatpox virus Pellor*, AY689436 *Deerpox virus W-848-83*, AY689437 *Deerpox virus W-1170-84*, MF966153 *White-tailed deer poxvirus isolate OV179*, MG751778 *Moosepox virus GoldyGopher14*) и геномы вирусов AJ293568 *Yaba-like disease virus YLDV*, EF420156 *Tanapox virus isolate TPV-Kenya*, HQ849551 *Yoka poxvirus strain DakArB 4268*. Кластер, в который входят геномы оспы подсемейства *Chordopoxvirinae* разных родов (AF170722 *Rabbit fibroma virus*, AF170726 *Myxoma virus strain Lausanne*, KT159937 *Salmon gill poxvirus*, KU980965 *Pteropox virus strain Australia*, MH427217 *Sea otter poxvirus strain ELK*, MN653921 *Cetacean poxvirus 1 strain CePV-TA*, MT712273 *Teiidae poxvirus 1 isolate 1642 19*). Кластер, включающий геномы оспы летучих мышей, дикобраза и мышей (HQ647181 *Cotia virus SPAn232*, KY747497 *Eptesipox virus strain Washington*, MK860688 *Hypsugopox virus strain 251170-23 2017*, MN692191 *Brazilian porcupinepox virus 1 strain UFU USP001*). Кластер, содержащий геномы оспы подсемейства *Chordopoxvirinae* с *GC*-составом больше 0.6 (AY386265 *Bovine papular stomatitis virus strain BV-AR02*, GQ329670 *Pseudocowpox virus strain VR634*, HE601899 *Squirrel poxvirus strain Red squirrel UK*, KM502564 *Parapoxvirus red deer HL953 strain HL953*, MN339351 *Equine molluscum contagiosum-like virus strain Tanzania 2016*, U60315 *Molluscum contagiosum virus subtype 1*). Пары близко

расположенных точек, соответствующие геномам оспы близкородственных видов (DQ356948 *Nile crocodilepox virus*, MG450915 *Saltwater crocodilepox virus subtype 1*), (MF467280 *Western grey kangaroopox virus strain Western Australia*, MF467281 *Eastern grey kangaroopox virus strain Sunshine Coast*) и (MF001304 *Murmansk poxvirus strain LEIV-11411*, MF001305 *NY 014 poxvirus strain 2013*). А также геномы оспы, точки, соответствующие которым, находятся достаточно удаленно от описанных выше кластеров (KR095315 *Diachasmimorpha longicaudata entomopoxvirus*, KY382358 *Seal parapoxvirus isolate AFK76s1*, LC613089 *Carp edema virus FTI2020 DNA*, KP728110 *Turkeypox virus strain TKPV-HU1124 2011*, MF503315 *Squirrelpox virus Berlin 2015*).

Приведенная выше кластеризация геномов оспы показывает, что наблюдается достаточно сильная зависимость между статистическими свойствами геномов оспы (в нашем случае условной энтропии) и таксономией носителей заболевания либо таксономией самого вируса. Особенно четко это видно на геномах оспы насекомых, оспы птиц и оспы парнокопытных. В то же время имеются кластеры, в которые попали геномы оспы одного рода *Orthopoxvirus* или одного подсемейства *Chordopoxvirinae*.

Финансирование: Базовый проект «Новые методы и технологии комплексного анализа сложных природных и антропогенных экосистем на основе современных средств моделирования и обработки данных, распределенных вычислений и цифрового мониторинга», FWES-2024-0014.

Relationship between information capacity and taxonomy of smallpox genomes

Senashova M.[1]*, Sadovsky M.[1, 2, 3]

[1] *Institute of Computational Modelling, SB RAS, Kasnoyarsk, Russia*

[2] *FSR&CC of FMBA of Russia, Krasnoyarsk, Russia*

[3] *Siberian Federal University, Krasnoyarsk, Russia*

* *msen@icm.krasn.ru*

Key words: nucleotide sequence; frequency; specific entropy; principal components; cluster

Motivation and Aim: Analysis of the statistical properties of nucleotide sequences is a core task of biophysics, bioinformatics and molecular biology. A number of papers have recently in this area (see, e.g. [1–4]). This paper examines the relationship between the information capacity of frequency dictionaries of nucleotide sequences with a thickness of 2 to 20 characters and the biological properties of these sequences. Each nucleotide sequence is assigned a point in 18-dimensional space. The coordinates of a point are the conditional entropies calculated for a given sequence. The composition of clusters formed by projections of points from 18-dimensional space into the space of the first three principal components is analyzed.

Methods and Algorithms: We will use frequency dictionaries it is a set of all fragments of a nucleotide sequence of a fixed length found in it, accompanied with their frequency [5–7]. An idea of the information capacity [8] characterizes sequences was studied as the conditional entropy of frequency dictionaries of thickness q, calculated from frequency dictionaries of smaller thickness (shorter words). The information capacity or

conditional entropy of the thickness dictionary q is provided by the formulae $\overline{S} = 2S_{q-1} - S_q - S_{q-2}$ or $\overline{S} = 2S_1 - S_2$ in case of $q = 2$; these are the absolute entropy of the thickness q of frequency dictionary. In our case $S_q = -\sum_{f_\omega} f_\omega \ln f_\omega$. Thus each genome is converted into a set of 18 figures of conditional entropy.

Results: 70 genomes of the family Poxviridae, from GenBank, have been examined. For these genomes, conditional entropies were calculated for dictionaries of thickness ranging from 1 to 20. The set of conditional entropies for each genome is considered as a point in 18-dimensional space. Next, the freely distributed VidaExpert software (http://bioinfo-out.curie.fr/projects/vidaexpert/) mapping the genomes into a set of points in principal components. We analyzed clusters consisting of points whose values of conditional entropy are close in the metric of Euclidean space. The following clusters were found: insect pox (AF063866 *Melanoplus sanguinipes entomopoxvirus O isolate Tucson*, AF250284 *Amsacta moorei entomopoxvirus*, AP013055 *Anomala cuprea entomopoxvirus DNA*, HF679131 *Adoxophyes honmai enomopoxvirus L*, HF679132 *Choristoneura biennis opoxvirus L*, HF679133 *Choristoneura rosaceana entomopoxvirus L*, HF679134 *Mythimna separata entomopoxvirus L*). The KR095315 *Diachasmimorpha longicaudata entomopoxvirus* genome did not fall into this cluster, since the *GC*-content of this genome is 0.3, and the *GC*-contents of the remaining insect pox genomes is about 0.2. Bird pox (AF198100 *Fowlpox virus*, AY318871 *Canarypox virus strain ATCC VR-111*, KJ801920 *Pigeonpox virus isolate FeP2*, KJ859677 *Penguinpox virus isolate PSan92*, MF678796 *Flamingopox virus FGPVKD09*, MK903864 *Magpiepox virus*, MT799800 *Che loniid poxvirus 1*, MW365933 *Albatrosspox virus strain SAN97-0665NZ*, MW485973 *Magpiepox virus 2 isolate 62-11-06-2000-ANU*). The same cluster included the turtlepox genome MT799800 *Cheloniid poxvirus 1*. A cluster including genomes of the genus *Orthopoxvirus* genomes (AF482758 *Cowpox virus strain Brighton Red*, AY009089 *Camelpox virus CMS*, DQ437594 *Taterapox* strain virus *Dahomey 1968*, HM172544 *Monkeypox virus strain Zaire 1979-005*, KP143769 *Raccoonpox virus*, KU749310 *Skunkpox virus strain WA*, KU749311 *Volepox virus strain CA*, L22579 *Variola major virus strain Bangladesh-1975*, MH607141 *Akhmeta virus isolate Akhmeta 2013-88*, MH816996 *Orthopoxvirus Abatino*, MN240300 *Borealpox virus*, OP526861 *Monkeypox virus isolate MPXV USA 2022 FL0019*, X69198 *Variola virus DNA*, *Y16780 Variola minor virus*). A cluster comprising the genomes of artiodactyl pox viruses (AF325528 *Lumpy skin disease virus NI-2490 isolate Neethling 2490*, AF410153 *Swinepox virus isolate 17077-99*, AY077832 *Sheeppox virus 10700-99 strain TU-V02127*, AY077835 *Goatpox virus Pellor*, AY689436 *Deerpox virus W -848-83*, AY689437 *Deerpox virus W-1170-84*, MF966153 *White-tailed deer poxvirus isolate OV179*, MG751778 *Moosepox virus GoldyGopher14*) and virus genomes AJ293568 *Yaba-like disease virus YLDV*, EF420156 *Tanapox virus isolate TPV-Kenya*, HQ849551 *Yoka poxvirus strain DakArB 4268*. Cluster comprising smallpox genomes of the subfamily *Chordopoxvirinae* of different genera (AF170722 *Rabbit fibroma virus*, AF170726 *Myxoma virus strain Lausanne*, KT159937 *Salmon gill poxvirus*, KU980965 *Pteropox virus strain Australia*, MH427217 *Sea otter poxvirus strain ELK*, MN653921 *Cetacean poxvirus 1 strain CePV-TA*, MT712273 *Teiidae poxvirus 1 isolate 1642 19*). A cluster comprising the smallpox genomes of bats, porcupines and mice (HQ647181 *Cotia virus SPAn232*, KY747497 *Eptesipox virus strain Washington*, MK860688 *Hypsugopox virus strain*

251170-23 2017, MN692191 *Brazilian porcupinepox virus 1 strain UFU USP001*). Cluster comprising smallpox genomes of the subfamily *Chordopoxvirinae* with *GC*-content greater than 0.6 (AY386265 *Bovine papular stomatitis virus strain BV-AR02*, GQ329670 *Pseudocowpox virus strain VR634*, HE601899 *Squirrel poxvirus strain Red squirrel UK*, KM502564 *Parapoxvirus red deer 3 strain HL953*, MN339351 *Equine molluscum contagiosum-like virus strain Tanzania 2016*, U60315 *Molluscum contagiosum virus subtype 1*). Pairs of closely located points corresponding to the smallpox genomes of closely related species: (DQ356948 *Nile crocodilepox virus*, MG450915 *Saltwater crocodilepox virus subtype 1*), (MF467280 *Western gray kangaroopox virus strain Western Australia*, MF467281 *Eastern gray kangaroopox strain virus Sunshine Coast*) and (MF001304 Murmansk poxvirus strain *LEIV-11411*, MF001305 *NY 014 poxvirus strain 2013*), as well as smallpox genomes, which corresponding points are located quite far from the clusters described above (KR095315 *Diachasmimorpha longicaudata entomopoxvirus*, KY382358 *Seal parapoxvirus isolate AFK76s1*, LC613089 *Carp edema virus FTI2020 DNA*, KP728110 *Turkeypox virus strain TKPV-HU1124 2 011*, MF503315 *Squirrelpox virus Berlin 2015*).

The above clustering of smallpox genomes shows that there is a fairly strong dependence between the statistical properties of smallpox genomes (in terms of conditional entropy) and the taxonomy of disease hosts, or the taxonomy of the virus itself. This is clearly are the genomes of insect pox, bird pox and artiodactyl pox. Simultaneously, there are clusters that contain smallpox genomes of one genus *Orthopoxvirus* or one subfamily *Chordopoxvirinae*.

Funding: Basic project “New methods and technologies for complex analysis of complex natural and anthropogenic ecosystems based on modern modeling and data processing tools, distributed computing and digital monitoring”, FWES-2024-0014.

Список литературы/References

1. Orlov Y.L., Boekhorst R.T., Abnizova I.A. Statistical measures of the structure of genomic sequences: Entropy, complexity, and position information. *J Bioinform Comput Biol*. 2006;4:523-536. doi 10.1142/S0219720006001801
2. Sabatti C., Rohlin L., Lange K., Liao J.C. Vocabulon: A dictionary model approach for reconstruction and localization of transcription factor binding sites. *Bioinformatics*. 2005;21:922-931. doi 10.1093/bioinformatics/bti083
3. Li W., Liu Y., Huang H.C. et al. Dynamical systems for discovering protein complexes and functional modules from biological networks. *IEEE/ACM Trans Comput Biol Bioinform*. 2007;4(2):233-250. doi 10.1109/TCBB.2007.070210
4. Гельфанд М.С. Апология биоинформатики. *Биофизика*. 2005;50(4):752-766
[Gelfand M.S. Apologia of bioinformatics. *Biophysics*. 2005;50(4):663-676 (in Russian)]
5. Bernaola-Galván P. et al. Study of statistical correlations in DNA sequences. *Gene*. 2002;300(1-2):105-115. doi 10.1016/s0378-1119(02)01037-5
6. Gorban A.N. et al. A new approach to the study of statistical properties of genetic sequences. *Biophysics*.1993;38(5):783-787
7. Горбань А.Н., Попова Т.Г., Садовский М.Г. Избыточность генетических текстов и мозаичная структура генома. *Молекулярная биология*. 1994;28(2):313-322
[Gorban A.N., Popova T.G., Sadovsky M.G. Redundancy of genetic sequences and mosaic structure of a genome. *Molekuliarnaia Biologiia*. 1994;28(2):313-322 (in Russian)]
8. Sadovsky M.G. Genomes and Information. *Biophysics*. 2009;54(4):419-422

doi 10.18699/bgrs2024-1.1-11

Деплеция нуклеопорина Elys не приводит к нарушению клеточного цикла в клетках S2 дрозофилы

Симонов Р.А., Михалева Е.А., Иванникова А.Я., Шевелев Ю.Я.*
Национальный исследовательский центр «Курчатовский институт», Москва, Россия
** shevelev@img.ras.ru*

Ключевые слова: нуклеопорин; Elys; кинетохор; клеточный цикл; дрозофила

Мотивация и цель: Компонент комплексов ядерных пор (КЯП) нуклеопорин Elys является ключевым фактором, определяющим сборку КЯП на выходе из митоза в многоклеточных организмах [1–4]. У млекопитающих, ксенопуса и нематоды во время митоза фракция Elys локализуется на кинетохорах [1, 2, 5, 6]. В соответствии с этим, деплеция Elys у этих организмов приводит к митотическим дефектам и задержке в прохождении клеточного цикла [1, 5, 6]. Однако в работе [7] было обнаружено, что в ходе митоза у дрозофилы Elys отсутствует на кинетохорах. В этой связи возникал вопрос, влияет ли деплеция Elys на клеточный цикл в клетках дрозофилы.
Методы и алгоритмы: 1.5×10^6 клеток S2 были собраны, дважды сполоснуты в PBS, зафиксированы холодным 70 % этанолом в течение 30 мин при 4 °C и промыты дважды в PBS. Клетки инкубировали 30 мин при 37 °C в 50 мкл PBS, содержащего 100 мкг/мл РНКазы А. Затем клетки инкубировали 10 мин при комнатной температуре в 250 мкл PBS, содержащего 40 мкг/мл пропидиум йодида. После этого с использованием проточного цитофлуориметра BD Accuri™ C6 Plus определяли долю клеток, находящихся на разных стадиях клеточного цикла, у примерно 10^4 контрольных клеток и клеток после нокдауна *Elys*. Перед процедурой цитофлуориметр калибровали на необработанных клетках S2.
Результаты: Сначала мы подтвердили, что Elys не локализуется на кинетохорах во время митоза у дрозофилы. Иммуноокрашивание клеток S2 антителами против Elys, против α-тубулина и против CenpA (последние антитела выявляют центромеры) показало, что во время метафазы Elys диффузно распределяется вокруг конденсированных хромосом, но отсутствует на кинетохорах. Затем мы провели проточный цитофлуориметрический анализ клеточного цикла в контрольных клетках S2 и в клетках S2 после нокдауна *Elys*. Эффективный нокдаун *Elys* в клетках S2 достигался методом РНК-интерференции. В результате мы обнаружили лишь незначительные изменения (если они вообще были) в прохождении клеточного цикла после нокдауна *Elys* в клетках S2 (рис. 1).
Выводы: Наши результаты показывают, что, в отличие от других организмов, в клетках S2 дрозофилы Elys не присутствует на кинетохорах во время митоза и его деплеция не приводит к нарушению клеточного цикла. Иными словами, существует четкая корреляция между наличием либо отсутствием Elys на кинетохорах в ряде организмов и аномальным либо нормальным митозом после деплеции Elys в этих организмах. На основании этого мы предполагаем, что Elys не влияет на митоз у дрозофилы по причине того, что он не связан с кинетохорами в этом организме.

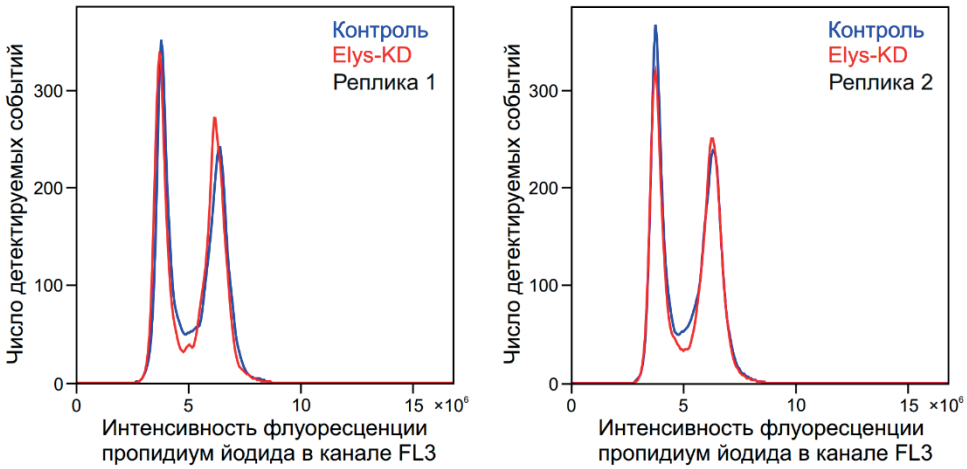


Рис. 1. Цитофлуориметрический анализ контрольных клеток S2 (синяя кривая) и клеток S2 после нокдауна Elys (красная кривая), ДНК которых была окрашена пропидиум йодидом. Показаны результаты для двух реплик. Пики, соответствующие клеткам до и после репликации, в контроле и опыте практически не отличаются

Финансирование: Исследование выполнено в рамках государственного задания НИЦ «Курчатовский институт».

Depletion of nucleoporin Elys does not result in cell cycle defects in *Drosophila* S2 cells

Simonov R.A., Mikhaleva E.A., Ivannikova A.Y., Shevelyov Y.Y.*

National Research Centre "Kurchatov Institute", Moscow, Russia

** shevelev@img.ras.ru*

Key words: nucleoporin; Elys; kinetochore; cell cycle; Drosophila

Motivation and Aim: Nucleoporin Elys, the constituent of nuclear pore complexes (NPCs), is a key factor determining NPC reassembly upon mitotic exit in multicellular organisms [1–4]. In mammals, *Xenopus* and nematode, a fraction of Elys is localized on kinetochores during mitosis [1, 2, 5, 6]. Accordingly, depletion of Elys in these organisms results in mitotic defects and in the delay of cell cycle progression [1, 5, 6]. However, Mehta et al. [7] revealed that, in *Drosophila*, Elys is absent on kinetochores during mitosis. Therefore, the question arises whether Elys depletion affects cell cycle progression in *Drosophila* cells.

Methods and Algorithms: Approximately 1.5×10^6 S2 cells were collected, rinsed twice in PBS, fixed with 70 % cold ethanol for 30 min at 4 °C and washed twice in PBS. Cells were incubated in 50 µl of PBS containing 100 µg/ml RNase A for 30 min at 37 °C. Then, cells were incubated in 250 µl of PBS containing 40 µg/ml propidium iodide for 10 min at room temperature. After that, the ratio of control and Elys-KD cells in different phases of cell cycle was analyzed for ~10^4 single cells using BD Accuri™ C6 Plus Flow Cytometer which has been calibrated on the untreated S2 cells.

Results: We first confirmed that Elys does not localize at kinetochores during mitosis in *Drosophila*. Immunostaining of S2 cells with anti-Elys, anti-α-tubulin and anti-CenpA

antibodies (the latter reveal the centromeres) has shown that, during metaphase, Elys is diffusely distributed around condensed chromosomes, and is absent on kinetochores. Next, we performed flow cytometry analysis of cell cycle progression in control S2 cells and in S2 cells after *Elys* knockdown. Efficient *Elys* knockdown in S2 cells was achieved by applying RNA-interference technique. As a result, we found only minor changes (if any) in cell cycle progression upon *Elys* knockdown in S2 cells (Fig. 1).

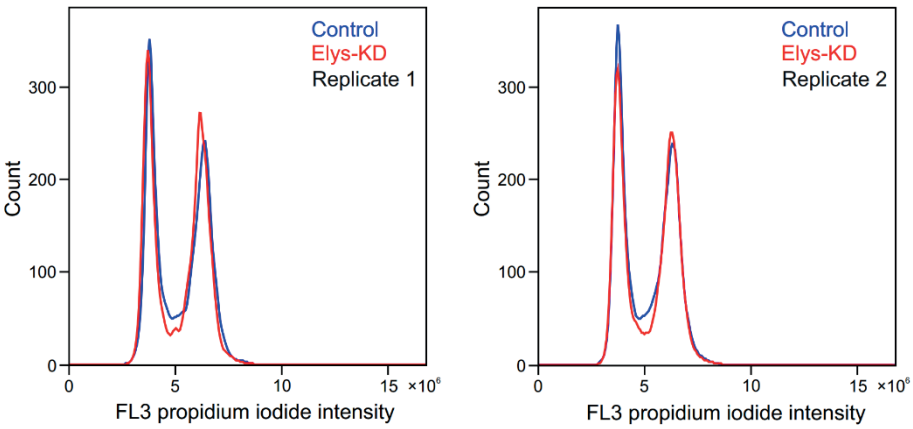


Fig. 1. Flow cytometry analysis of control (blue curve) and Elys-KD (red curve) S2 cells after DNA staining with propidium iodide. Two replicates are indicated. The peaks corresponding to cells before and after replication are practically the same in the control and after Elys-KD

Conclusion: Our results indicate that, unlike in other organisms, in *Drosophila* S2 cells Elys does not stay at kinetochores during mitosis and its depletion does not lead to the disturbance of cell cycle progression. In other words, there is a clear correlation between Elys presence or absence on kinetochores in the organisms and abnormal or normal mitosis upon Elys depletion in these organisms. We, therefore, suggest that *Drosophila* Elys does not affect mitosis because it is not bound with kinetochores in this organism.

Funding: The study was carried out within the framework of the state task of NRC "Kurchatov Institute".

Список литературы/References

1. Rasala B.A., Orjalo A.V., Shen Z., Briggs S., Forbes D.J. ELYS is a dual nucleoporin/kinetochore protein required for nuclear pore assembly and proper cell division. *Proc Natl Acad Sci USA.* 2006;103:17801-17806. doi 10.1073/pnas.0608484103
2. Franz C., Walczak R., Yavuz S., Santarella R., Gentzel M., Askjaer P., Galy V., Hetzer M., Mattaj I.W., Antonin W. MEL-28/ELYS is required for the recruitment of nucleoporins to chromatin and postmitotic nuclear pore complex assembly. *EMBO Rep.* 2007;8:165-172. doi 10.1038/sj.embor.7400889
3. Gillespie P.J., Khoudoli G.A., Stewart G., Swedlow J.R., Blow J.J. ELYS/MEL-28 chromatin association coordinates nuclear pore complex assembly and replication licensing. *Curr Biol.* 2007;17:1657-1662. doi 10.1016/j.cub.2007.08.041
4. Rasala B.A., Ramos C., Harel A., Forbes D.J. Capture of AT-rich chromatin by ELYS recruits POM121 and NDC1 to initiate nuclear pore assembly. *Mol Biol Cell.* 2008;19:3982-3996. doi 10.1091/mbc.e08-01-0012
5. Fernandez A.G., Piano F. MEL-28 is downstream of the Ran cycle and is required for nuclear-envelope function and chromatin maintenance. *Curr Biol.* 2006;16:1757-1763. doi 10.1016/j.cub.2006.07.071
6. Galy V., Askjaer P., Franz C., López-Iglesias C., Mattaj I.W. MEL-28, a novel nuclear-envelope and kinetochore protein essential for zygotic nuclear-envelope assembly in C. elegans. *Curr Biol.* 2006;16:1748-1756. doi 10.1016/j.cub.2006.06.067
7. Mehta S.J.K., Kumar V., Mishra R.K. Drosophila ELYS regulates Dorsal dynamics during development. *J Biol Chem.* 2020;295:2421-2437. doi 10.1074/jbc.RA119.009451

doi 10.18699/bgrs2024-1.1-12

Эволюционно консервативные по синтении гены и нейродегенеративные патологии

Скобель О.И.*, Глазко В.И., Глазко Т.Т., Косовский Г.Ю.
Научно-исследовательский институт пушного звероводства и кролиководства им. В.А. Афанасьева, Московская область, Раменский район, Россия
** skobelolga@gmail.com*

Ключевые слова: *kcne2*; *gart*; *tmem50b*; *il10rb*; *ifnar2*; *urb1*; *grik1*; *usp16*; *ltn1*; *cyyr1*; *app*; *jam2*; синтения; микроРНК; регуляторные сети

Мотивация и цель: Изучение консервативности межвидовой синтении является важной основой для проведения геномных исследований, необходимо для понимания функциональных особенностей, входящих в консервативный синтенный блок генов, а также широко используется для анализа путей эволюционной дивергенции различных организмов [1, 2]. Ранее в наших исследованиях выявлено 12 генов (*kcne2*, *gart*, *tmem50b*, *il10rb*, *ifnar2*, *urb1*, *grik1*, *usp16*, *ltn1*, *cyyr1*, *app*, *jam2*), несущих продукт рекомбинации между двумя мобильными элементами (BovB и ERV) с участками гомологии к различным микроРНК и характеризующимися высокой эволюционной консервативностью синтении у млекопитающих, начиная с утконоса [3]. У человека блок этих генов локализован в хромосоме 21, трисомия которой приводит к нейродегенеративному заболеванию – синдрому Дауна. Для того чтобы выяснить возможные функциональные связи, ассоциированные с эволюционной консервативностью синтении этих генов, выполнен анализ *in silico* специфических функциональных особенностей их экспрессии.
Методы и алгоритмы: Информация об исследуемых структурных генах и их локализации в геномах различных видов получена из международной базы данных GenBank. Для построения сети известного и предсказанного взаимодействия использовали ресурсы GeneMANIA, а также STRING ver. 12.0. Анализ функциональной аннотации выполнен с помощью веб-сервиса DAVID, с использованием GOA, InterPro, KEGG, SMART, UniProt. Проанализированы литературные источники, индексируемые в международных базах данных Web of Science, Scopus, Pubmed.
Результаты: Изучаемые структурные гены являются частью одной сети, чья биологическая значимость нуждается в дальнейшем исследовании. Оказалось, что их экспрессия не тканеспецифична. Белки, кодируемые этими генами, участвуют в фундаментальных биологических процессах жизнеобеспечения клеток, таких как пуриновый синтез, белковый синтез, межклеточные взаимодействия, ионная возбудимость клеток (калиевые каналы), воспалительные и антиоксидантные процессы. Перечисленные функции полигенны, соответствующие геномные элементы разбросаны по разным хромосомам, повышенная или пониженная индивидуальная экспрессия отдельных генов из блока 12 синтенных генов обнаруживается в ряде опухолей, например, молочной железы, печени. Тем не менее изменения экспрессии последних наблюдается у всех 12 генов при таких нейродегенеративных патологиях, как болезнь Альцгеймера и синдром Дауна [4]. Полученные результаты подтверждают имеющиеся данные о том, что в эволю-

ционно консервативных по генетическому сцеплению блоках повышена плотность генов, вовлеченных в формирование анатомических характеристик и развитие центральной нервной системы [5]. С функциональной точки зрения эти гены подразделяются на следующие группы: угнетение клеточного деления и клеточного обновления, пониженная возбудимость клеток, повышение риска воспалительных реакций и окислительного стресса, нарушения межклеточных взаимодействий. Тесная функциональная связь выявлена между нарушением возбудимости клеток и белка предшественника амилоида (APP и белок калиевого канала – KCNE2; APP и уменьшение утилизации глутамата и увеличение окислительного стресса – GRIK1; изменения активности путей синтеза пуринов – GART; выброс цитокинов – IFNAR2 и IL10RB), ускорения старения клеток (USP16), дефектов белкового синтеза (проверка качества рибосом – LTN1 и дефектные рибосомы – URB1), межклеточных взаимодействий (APP; плотные межклеточные контакты – JAM2; транспортировка и формирование эндо- и экзосом – CYYR1 и TMEM50B).

Выводы: По-видимому, именно взаимосвязями между нарушением межклеточных взаимодействий, старением и деградацией клеток может объясняться эволюционная консервативность синтении рассмотренных генов и присутствие в них одинаковых элементов регуляторной сети – определенных микроРНК, вносящих возможность синхронизированной коррекции их экспрессии.

Финансирование: Исследование поддержано Министерством науки и высшего образования Российской Федерации (№ 075-00503-24-01).

Evolutionarily conserved genes by synteny and neurodegenerative pathologies

Skobel O.I.*, Glazko V.I., Glazko T.T., Kosovsky G.Yu.

Afanas`ev Research Institute of Fur – Bearing Animal Breeding and Rabbit Breeding, Moscow region, Ramensky district, Russia

** skobelolga@gmail.com*

Key words: *kcne2*; *gart*; *tmem50b*; *il10rb*; *ifnar2*; *urb1*; *grik1*; *usp16*; *ltn1*; *cyyr1*; *app*; *jam2*; synteny; microRNA; regulatory networks

Motivation and Aim: In addition to serving as a crucial foundation for genomic research, the study of interspecific synteny conservation is also extensively used to examine the evolutionary divergence pathways in various organisms and to comprehend the functional characteristics of the genes contained in the conserved synteny block [1, 2]. In our previous studies, we identified twelve genes – *kcne2*, *gart*, *tmem50b*, *il10rb*, *ifnar2*, *urb1*, *grik1*, *usp16*, *ltn1*, *cyyr1*, *app*, and *jam2* – carrying a recombination product between two transposons – BovB and ERV – with homology sites to distinct microRNAs [3]. These genes form a highly evolutionary-conserved synteny block in mammals. This block is located on chromosome 21 in Homo sapiens, the trisomy of which causes neurodegenerative diseases such as Down syndrome. Analyzing *in silico* the distinct functional aspects of these genes' expression allowed us to piece together their potential functional connections that could explain the evolutionary conservatism of their synteny.

Methods and Algorithms: Information on the studied structural genes and their localization in the genomes of different species was obtained from the GenBank

database. To construct the network of known and predicted interactions, we used the GeneMANIA prediction server as well as STRING (version 12.0). Functional annotation analysis was performed with the DAVID resource system using GOA, InterPro, KEGG, SMART, and UniProt. Literature sources from the Web of Science, Scopus, and Pubmed were analyzed.

Results: The structural genes studied are part of the same network, the biological significance of which needs to be further investigated. It has been shown that their expression is not tissue-specific. These genes encode proteins that are involved in fundamental biological processes of cell life support, such as purine synthesis, protein synthesis, intercellular interactions, ionic excitability of cells (potassium channels), and inflammatory and antioxidant processes. The listed functions are polygenic, and the associated genomic elements are dispersed across different chromosomes. Increased or decreased individual expression of individual genes from the block of 12 syntenic genes is found in a number of tumors, e.g., breast and liver. However, changes in the expression of all 12 genes are observed in neurodegenerative diseases such as Alzheimer's disease and Down syndrome [4]. The results obtained support existing data that the density of genes involved in anatomical and central nervous system development is increased in evolutionarily conserved genetic linkage blocks [5]. Functionally, these genes are classified into the following groups: inhibition of cell division and renewal, reduced cell excitability, increased risk of inflammation and oxidative stress, and impaired intercellular interactions. A close functional relationship has been found between impaired cell excitability and amyloid precursor protein (APP and potassium channel protein – KCNE2; APP and decreased glutamate utilization and increased oxidative stress – GRIK1; changes in purine synthesis pathway activity – GART; cytokine release – IFNAR2 and IL10RB), accelerated cell aging (USP16), defects in protein synthesis (ribosome quality control – LTN1 and defective ribosomes – URB1), intercellular interactions (APP; tight intercellular contacts – JAM2; transport and formation of endo- and exosomes – CYYR1 and TMEM50B).

Conclusion: Apparently, it is the link between the disruption of intercellular interactions, aging, and cell degradation that can explain the evolutionary conservatism of the synteny of the genes studied and the presence in them of the same elements of the regulatory network – certain microRNAs that allow the synchronized correction of their expression.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (No. 075-00503-24-01).

Список литературы/References

1. Ghiurcuta C.G., Moret B.M.E. Evaluating synteny for improved comparative studies. *Bioinformatics.* 2014;30(12):i9-i18. doi 10.1093/bioinformatics/btu259
2. Simakov O., Marlétaz F., Yue J.X. et al. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat Ecol Evol*. 2020;4(6):820-830. doi 10.1038/s41559-020-1156-z
3. Скобель В.И., Глазко Г.Ю., Косовский Т.Т., Глазко О.И. Рекомбинации между мобильными генетическими элементами как источник микроРНК *Известия Тимирязевской сельскохозяйственной академии*. 2017;4:70-98. doi 10.26897/0021-342X-2017-4-70-98
[Skobel O.I., Glazko V.I., Kosovsky G.Yu., Glazko T.T. Recombinations among mobile genetic elements as a source of microRNA. *Izv Timirâzevsk s-h akad*. 2017;4:70-98 (in Russian)]
4. Vilardell M., Rasche A., Thormann A. et al. Meta-analysis of heterogeneous Down Syndrome data reveals consistent genome-wide dosage effects related to neurological processes. *BMC Genomics*. 2011;12(1):229. doi:10.1186/1471-2164-12-229
5. Damas J., Corbo M., Kim J. et al. Evolution of the ancestral mammalian karyotype and syntenic regions. *Proc Natl Acad Sci USA*. 2022;119(40):e2209139119. doi 10.1073/pnas.2209139119

doi 10.18699/bgrs2024-1.1-13

Различия в организации структуры цис-регуляторных областей генов человека в зависимости от их тканеспецифичности

Слюсарев С.С.*, Бутенко Е.В., Шкурат Т.П.
Южный федеральный университет, Ростов-на-Дону, Россия
** slusarev-sergey@yandex.ru*

Ключевые слова: тканеспецифичные гены; CpG островки; Alu-элементы; организация генома

Мотивация и цель: Гены человека можно разделить на повсеместно экспрессируемые гены домашнего хозяйства и тканеспецифичные гены, которые экспрессируются только в одном типе клеток или тканей. Генетический механизм регуляции этих типов генов может быть ассоциирован с организацией цис-регуляторного региона. CpG-островки, расположенные в цис-регуляторных областях, могут изменять уровни экспрессии генов в зависимости от паттерна их метилирования.
Целью данной работы был поиск структурных отличий функциональных элементов в цис-регуляторных регионах тканеспецифичных генов от генов повсеместно экспрессирующихся (генов домашнего хозяйства).
Методы и алгоритмы: Материалом исследования послужили 977 тканеспецифичных генов человека из базы Human Protein Atlas и 554 гена «домашнего хозяйства» [1]. Анализ функциональных элементов проводили с помощью Cpgplot, NCBI Genome browser, SINE Base, а также баз данных об экспрессии генов GTEx, Illumina, BioGPS. Исследовали наличие CpG-островков в цис-регуляторных областях (2000 п. н. до сайта начала транскрипции) изучаемых генов.
Результаты: В TSS200 области генов домашнего хозяйства в 88.08 % были представлены GC островками, в TSS1500 области они составляли только 31.04 % от общего числа последовательностей. Для тканеспецифичных генов было обнаружено значительно меньшее число GC островков в TSS200 области (19.55 %), и 9.31 % этих генов имели GC островки в TSS1500 области. Отмечается существенная разница и в длине островков между исследуемыми типами генов. Обнаружено, что средняя длина островка равнялась 566.51 п. н. для тканеспецифичных генов и 719.71 п. н для генов домашнего хозяйства. С помощью базы данных SINE Base осуществляли поиск гомологии для коротких диспергированных повторов в цис-регуляторных областях изучаемых генов. Alu-элементы присутствовали в 49.84 % цис-регуляторных областей тканеспецифичных генов и в 65.52 % цис-регуляторных областей генов домашнего хозяйства. Обнаружено, что усредненное расстояние от Alu-элемента до сайта начала транскрипции генов домашнего хозяйства меньше, чем это же расстояние для тканеспецифичных генов. Биоинформатический анализ показал, что пары Alu-элементов присутствовали в 13.2 и 23.1 % цис-регуляторных областей тканеспецифичных генов и генов домашнего хозяйства соответственно. Важно отметить, что все Alu-пары имели инвертированный характер ориентации. Среднее расстояние между инвертированными парами Alu-элементов, найденных в цис-регуляторных

областях тканеспецифичных генов было меньше, чем аналогичный показатель для генов домашнего хозяйства.
Выводы: Обнаружены существенные отличия в локализации CpG-островков и Alu-элементов цис-регуляторных регионов тканеспецифичных генов от генов повсеместно экспрессирующихся (генов домашнего хозяйства) в TSS200 и TSSS1500.
Финансирование: Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках государственного задания в сфере научной деятельности № FENW-2023-0018.

Differences in the organisation of the structure of cis-regulatory regions of human genes depending on their tissue specificity

Slyusarev S.S.*, Butenko E.V., Shkurat T.P.
Southern Federal University, Rostov-on-Don, Russia
** slusarev-sergey@yandex.ru*

Key words: tissue-specific genes, CpG islands, Alu elements, genome organization

Motivation and Aim: Human genes can be divided into ubiquitously expressed housekeeping genes and tissue-specific genes that are expressed in only one type of cell or tissue. The genetic mechanism of regulation of these types of genes may be associated with the organisation of the cis-regulatory region. CpG islands located in cis-regulatory regions may change the expression levels of genes depending on their methylation pattern. The aim of this work was to find structural differences between the functional elements in the cis-regulatory regions of non-specific genes and genes that are universally expressed (housekeeping genes).
Methods and Algorithms: The study material included 977 human tissue-specific genes from the Human Protein Atlas database and 554 housekeeping genes (Eisenberg E., Levanon E. Y., 2013). Analysis of functional elements was carried out using Cpgplot, NCBI Genomebrowser, SINE Base, as well as gene expression databases GTEx, Illumina, BioGPS. The presence of CpG islands in cis-regulatory regions (2000 bp before the transcription start site) of the studied genes was studied.
Results: 88.08 % of housekeeping genes contained CpG islands in the TSS200 region, and only 31.04 % of ubiquitously expressed genes contained a CpG island in the TSS1500 region. For tissue-specific genes a significantly smaller number of CpG islands were found in the studied regions. 19.55 % of tissue-specific genes contained CpG islands in the TSS200 region, and only 9.31% of these genes contained CpG islands in the TSS1500 region. There is also a significant difference in the length of islands between the types of genes studied. It was found that the average length of the island was 566.51 bp. for tissue-specific genes and 719.71 bp for housekeeping genes. Using the SINE Base database, we searched for homology for short dispersed repeats in the cis-regulatory regions of the studied genes. Alu elements were present in 49.84 % of the cis-regulatory regions of tissue-specific genes and in 65.52 % of the cis-regulatory regions of housekeeping genes. It was found that the average distance from the Alu element to the transcription start site of home genes is less than the same distance for tissue-specific genes. Bioinformatics analysis showed that Alu element pairs were

present in 13.2 and 23.1 % of cis-regulatory regions of tissue-specific and housekeeping genes, respectively. It is important to note that all Alu pairs had an inverted orientation. The average distance between inverted pairs of Alu elements found in cis-regulatory regions of tissue-specific genes was less than that for housekeeping genes.

Conclusion: Significant differences were found in the localization of CpG islands and Alu elements of cis-regulatory regions in tissue-specific genes from genes ubiquitously expressed in the TSS200 and TSSS1500 regions.

Funding: The work was carried out with financial support from the Ministry of Science and Higher Education of the Russian Federation within the framework of the state assignment in the field of scientific activity No. FENW-2023-0018.

Список литературы/References

1. Eisenberg E., Levanon E.Y. Human housekeeping genes, revisited. *Trends Genet*. 2013;29(10):569-574

doi 10.18699/bgrs2024-1.1-14

Генетическое профилирование штаммов *Ligilactobacillus salivarius*, устойчивых к желчным кислотам

Хуснутдинова Д.[1]*, Маркелова М.[1], Синягина М.[1], Сенина А.[1], Булыгина Е.[1], Абдулхаков С.[1, 2], Григорьева Т.[1]

[1] *Казанский (Приволжский) федеральный университет, Казань, Россия*

[2] *Казанский государственный медицинский университет, Казань, Россия*

* *dilyahusn@gmail.com*

Ключевые слова: *Ligilactobacillus salivarius*; желчные кислоты; болезнь Крона; полногеномное секвенирование

Мотивация и цель: Микробиота кишечника принимает непосредственное участие в биотрансформации желчных кислот с помощью микробных ферментов. Одним из этапов метаболизма желчных кислот является деконъюгация первичных желчных кислот с помощью бактериальных гидролаз желчных кислот (BSH). Лактобациллы, в частности *L. salivarius*, принимают активное участие в данных процессах и обладают большим количеством генов глицин- и таурин-гидролаз [1, 2]. Кишечник пациентов с болезнью Крона, помимо характерного дисбиоза, имеет в составе большое количество желчных кислот, в особенности коньюгированных [3–5]. В связи с этим целью работы было сравнить желчеустойчивость и генетический потенциал штаммов *L. salivarius*, выделенных от пациентов с болезнью Крона и здоровых добровольцев.

Методы и алгоритмы: Объектом исследования были штаммы *L. salivarius*: 3 – выделенные из микробиоты кишечника пациентов с болезнью Крона, 4 – выделенные от здоровых добровольцев и 1 – из пробиотического препарата OMNI-biotic (AllergoSan, Германия). Для оценки желчеустойчивости культуры выращивали при 37 °C на среде Man–Rogosa–Sharpe (MRS) с бычьей желчью в концентрациях: 0.3 %, 0.5 %, 1.0 %, 1.5 %, 5.0 %, 7.5 %. Устойчивость оценивали по изменению оптической плотности бактериальной культуры через 4, 24 и 48 часов. Выделение геномной ДНК проводили с использованием набора ZymoBIOMICS DNA Miniprep Kit (Zymo Research, США) с модификациями. Подготовку библиотек ДНК осуществляли с использованием набора NEBNext Ultra II для Illumina (New England BioLabs, США) согласно рекомендациям производителя. Секвенирование проводили на платформе Illumina MiSeq (в режиме 2*300 п.н.).

Результаты: Штаммы *L. salivarius*, выделенные от пациентов с болезнью Крона, не отличаются по желчеустойчивости от штаммов здоровых добровольцев и пробиотического штамма при выбранных концентрациях (до 7.5 % желчи).

Согласно результатам полногеномного анализа, геномы штаммов *L. salivarius*, выделенные от пациентов с болезнью Крона и от здоровых добровольцев, имеют один общий ген *bsh*, кодирующий холоилглицингидролазу – WP_014568264.1 (табл. 1), однако по второму гомогичному гену *bsh* штаммы отличаются. Штаммы, выделенные от здоровых добровольцев, обладают одним общим мотивом белка, а штаммы от пациентов с болезнью Крона – другим, WP_003699092.1.

Пробиотический штамм *L. salivarius Di6* обладает двумя гомологичными генами *bsh* (WP_034983830.1; WP_032494132.1), однако они отличаются от генов в штаммах, выделенных от участников исследования.

Таблица 1. Гены *bsh,* обнаруженные в штаммах *L. salivarius*

Группа	Штамм	Продукт	№ RefSeq
Здоровые добровольцы	A5-1-10	белок семейства холоилглицингидролаз	WP_014568264.1
		холоилглицингидролаза	protein motif:HMM:NF038245.1“ /note="Derived by automated computational analysis using gene prediction method: Protein Homology"
	A5-2-10	белок семейства холоилглицингидролаз	WP_014568264.1
		холоилглицингидролаза	protein motif:HMM:NF038245.1"“ /note="Derived by automated computational analysis using gene prediction method: Protein Homology"
	A7-1-15	белок семейства холоилглицингидролаз	WP_014568264.1
		холоилглицингидролаза	protein motif:HMM:NF038245.1"“ /note="Derived by automated computational analysis using gene prediction method: Protein Homology"
	A7-2-5	белок семейства холоилглицингидролаз	WP_014568264.1
		холоилглицингидролаза	protein motif:HMM:NF038245.1"“ /note="Derived by automated computational analysis using gene prediction method: Protein Homology"
Пациенты с болезнью Крона	198vzk-4	белок семейства холоилглицингидролаз	WP_014568264.1
		холоилглицингидролаза	WP_003699092.1
	144-2021-8	холоилглицингидролаза	WP_003699092.1
		белок семейства холоилглицингидролаз	WP_014568264.1
	098-2021-6	белок семейства холоилглицингидролаз	WP_014568264.1
		холоилглицингидролаза	WP_003699092.1
Пробиотический штамм	Di6	холоилглицингидролаза	WP_034983830.1
		холоилглицингидролаза	WP_032494132.1

Выводы: Обнаружены различия в структуре генов *bsh* – гидролаз желчных кислот, которые не связаны с устойчивостью *L. salivarius* к желчным кислотам. Вероятно, данные различия могут обеспечивать субстратную специфичность к глицин- и/или таурин-конъюгированным желчным кислотам.

Финансирование: Работа выполнена на базе Междисциплинарного центра коллективного пользования КФУ в рамках Программы стратегического академического лидерства Казанского федерального университета (ПРИОРИТЕТ-2030) и финансируется за счет средств субсидии, выделенной Казанскому федеральному университету на выполнение государственного задания в сфере научной деятельности (проект № FZSM-2023-0013).

Genetic profiling of strains of *Ligilactobacillus salivarius* resistant to bile acid

Khusnutdinova D.[1]*, Markelova M.[1], Siniagina M.[1], Senina A.[1], Boulygina E.[1], Abdulkhakov S.[1, 2], Grigoryeva T.[1]

[1] *Kazan Federal University, Kazan, Russia*

[2] *Kazan State Medical University, Kazan, Russia*

* *dilyahusn@gmail.com*

Key words: *Ligilactobacillus salivarius*; bile acids; Crohn's disease; whole genome sequencing

Motivation and Aim: The gut microbiota is directly involved in the biotransformation of bile acids through microbial enzymes. One step in bile acid metabolism is the deconjugation of primary bile acids by bacterial bile acid hydrolases (BSH). *Lactobacillaceae*, particularly *L. salivarius*, play an active role in these processes and have a large number of glycine and taurine hydrolase genes [1, 2]. In addition to the specific dysbiosis, the intestines of patients with Crohn's disease contain high levels of bile acids, particularly conjugated bile acids [3–5]. The aim of the study was to compare the resistance to bile in various concentrations and to assess genetic potential of *L. salivarius* strains isolated from patients with Crohn's disease and healthy volunteers.
Methods and Algorithms: The objects of the study were *L. salivarius* strains: 3 – isolated from the intestinal microbiota of patients with Crohn's disease, 4 – isolated from healthy volunteers and 1 – from probiotic OMNI-biotic (AllergoSan, Germany). To assess bile resistance, bacterial cultures were grown at 37 °C on Man–Rogosa–Sharpe (MRS) medium with bovine bile in concentrations: 0.3 %; 0.5 %; 1.0 %; 1.5 %; 5.0 %; 7.5 %. The Resistance was determined according to changes in optical density after 4, 24 and 48 hours of culturing. Genomic DNA was extracted from strains using the ZymoBIOMICS DNA Miniprep Kit (Zymo Research, USA) with modifications. DNA libraries were prepared using the NEBNext Ultra II kit for Illumina (New England BioLabs, USA). Sequencing was performed on an Illumina MiSeq platform (2*300 bp mode).
Results: *L. salivarius* strains isolated from patients with Crohn's disease do not differ in bile resistance from strains from healthy patients and from the probiotic strain at tested concentrations (up to 7.5 % bile).
According to genome-wide analysis, the genomes of *L. salivarius* strains isolated from both Crohn's disease patients and healthy volunteers have the same *bsh* gene encoding choloylglycine hydrolase – WP_014568264.1 (Table 1), but the strains differ in the second homologous *bsh* gene. The Strains isolated from healthy volunteers have one common protein motif and the strains from Crohn's disease patients have another one – WP_003699092.1.

The probiotic strain *L. salivarius Di6* has two homologous *bsh* genes (WP_034983830.1; WP_032494132.1), but they differ from the genes in the strains isolated from the study participants.

Table 1. *bsh* genes found in *L. salivarius* strains

Group	Strain	Product	№ RefSeq
Healthy volunteers	A5-1-10	choloylglycine hydrolase family protein	WP_014568264.1
		choloylglycine hydrolase	protein motif:HMM:NF038245.1“ /note="Derived by automated computational analysis using gene prediction method: Protein Homology"
	A5-2-10	choloylglycine hydrolase family protein	WP_014568264.1
		choloylglycine hydrolase	protein motif:HMM:NF038245.1"“ /note="Derived by automated computational analysis using gene prediction method: Protein Homology"
	A7-1-15	choloylglycine hydrolase family protein	WP_014568264.1
		choloylglycine hydrolase	protein motif:HMM:NF038245.1"“ /note="Derived by automated computational analysis using gene prediction method: Protein Homology"
	A7-2-5	choloylglycine hydrolase family protein	WP_014568264.1
		choloylglycine hydrolase	protein motif:HMM:NF038245.1"“ /note="Derived by automated computational analysis using gene prediction method: Protein Homology"
Crohn's disease patients	198vzk-4	choloylglycine hydrolase family protein	WP_014568264.1
		choloylglycine hydrolase	WP_003699092.1
	144-2021-8	choloylglycine hydrolase	WP_003699092.1
		choloylglycine hydrolase family protein	WP_014568264.1
	098-2021-6	choloylglycine hydrolase family protein	WP_014568264.1
		choloylglycine hydrolase	WP_003699092.1
Probiotic	Di6	choloylglycine hydrolase	WP_034983830.1
		choloylglycine hydrolase	WP_032494132.1

Conclusion: We found the differences in the structure of the *bsh* genes – bile acid hydrolases which are not related to the resistance of *L. salivarius* to bile acids. It is likely that these differences confer a substrate specificity for glycine and/or taurine conjugated bile acids.

Funding: The work was carried out on the basis of the Interdisciplinary Center for Collective Use of Kazan Federal University within the framework of the Strategic Academic Leadership Program of the Kazan Federal University (PRIORITY-2030) and was financed by a subsidy allocated to the Kazan Federal University for the implementation of the state assignment in the field of scientific activity (project No. FZSM-2023-0013).

Список литературы/References

1. Babu A.F. et al. Identification and distribution of sterols, bile acids, and acylcarnitines by LC-MS/MS in humans, mice, and pigs-a qualitative analysis. *Metabolites*. 2022;12(1):49
2. Li C.Y. et al. Discovery and characterization of naturally occurring chalcones as potent inhibitors of bile salt hydrolases. *Acta Mater Med.* 2022;1(2):169-176
3. Gadaleta R.M. et al. Bile salt hydrolase-competent probiotics in the management of IBD: unlocking the “Bile acid code”. *Nutrients*. 2022;14(15):3212
4. Fitzpatrick L.R., Jenabzadeh P. IBD and bile acid absorption: focus on pre-clinical and clinical observations. *Front Physiol*. 2020;11:564
5. Thomas J.P. et al. The emerging role of bile acids in the pathogenesis of inflammatory bowel disease. *Front Immunol*. 2022;13:246

doi 10.18699/bgrs2024-1.1-15

Prediction of operon-like gene clusters in the *Intoshia linei* genome

Bondarenko N.I.
Department of Invertebrate Zoology, Faculty of Biology, St. Petersburg State University, St. Petersburg, Russia
Natalya.lannik@gmail.com

Key words: Orhtonectida; genome reduction; operon; adaptations

Motivation and Aim: Orthonectida is a small, rare, and in many aspects enigmatic group of organisms with a unique life cycle and a highly simplified adult free-living stage parasitizing various marine invertebrates [1, 2]. Several studies have shown that parasitism can lead to a dramatic reduction of the body plan, morphological structures and also affects organisms at the genomic level [3, 4]. In previous studies were shown that Orthonectida has extremely compact genomes due to the significant reduction of gene number, intergenic regions, intron length, and repetitive elements [5, 6]. As a result, gene density in these species much higher than in other metazoans. For example, in *Intoshia linei* genes cover 53.8 % of the genome, an average density of 202 genes per Mb and median intergenic distance consists of 1549 bp [6]. The close proximity of Orthonectida genes suggested that they might be possess operon-like gene organization and co-transcribed from a single upstream promoter. In this study we sequenced *I. linei* transcriptome and compared the experimentally determined polycistronic transcripts with computational analysis of intergenic distances in the *I. linei* genome.
Methods and Algorithms: To identify potential orthonectid operon-like gene clusters we employed our pipeline with the prediction method based on the intergenic distance between reading frames as a primary predictor of the number of operons in the genome. All predicted operon-like gene clusters then were validated using the RNA-seq data. Total RNA was isolated using Quick-RNA Miniprep (Zimo Research). The RNA-seq library was prepared using NEBNext RNA library preparation protocol (New England BioLabs). The library was sequenced on a HiSeq 2000 sequencing system, producing 52.2 M 110-bp paired-end reads. The reads were adapter trimmed with Trimmomatic v0.30 and assembled into transcripts using the rnaSPAdes. To find polycistronic transcripts we used TransDecoder and our own scripts for excluding false positive prediction.
Results: Computational analysis of the genome assembly for *I. linei* predicted that 4497 genes (51.5 %) have intergenic distance ≤ 3500 bp and might be organized into 1847 operons. The average number of genes per operon are 2.4, minimum two genes per operon, and maximum 9 genes per operon. The median intergenic distance within the set of all these predicted operons in *I. linei* was observed to be 1285 nt, while 934 operons (~50.5 % of the total 1847 operons) have genes with intergenic distances < 500 nt. According to this, most of the predicted *I. linei* operons are characterized by close apposition of the cotranscribed genes. To characterize of *I. linei* transcriptomes and investigate the presence of operon-derived mRNAs total RNA-sequencing was used. Sequencing revealed 403 dicistronic and 55 polycistronic transcripts. More than 4 % of the transcriptome of *I. linei* is organized as operons. Thereby we validated 982 genes

organized into operons out of 4497 potential candidates. Most of the genes in *I. linei* operons are separated from one another by approximately 1289 bp.

Conclusion: In this study we have predicted and validated that *I. linei* genome has operon-like nonoverlapping gene clusters which are transcribed as polycistronic transcripts.

Funding: The study is supported with Russian Science Foundation 24-24-00449.

References

1. Kozloff E.N. The genera of the phylum Orthonectida. *Cah Biol Mar*. 1992;33:377-406
2. Slyusarev G.S. Phylum Orthonectida: morphology, biology, and relationships to other multicellular animals. *Zh Obshch Biol*. 2008;69:403-427
3. Burke M. et al. The plant parasite *Pratylenchus coffeae* carries a minimal nematode genome. *Nematology*. 2017;(6):621-637. doi 10.1163/15685411-00002901
4. Wang J., Gao S., Mostovoy Y., Kang Y. et al. Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Res*. 2017;27(12):2001-2014. doi 10.1101/gr.225730.117
5. Mikhailov K.V., Slyusarev G.S., Nikitin M. et al. The genome of *Intoshia linei* affirms orthonectids as highly simplified spiralians. *Curr Biol*. 2016;26:1768-1774
6. Slyusarev G., Starunov V., Bondarenko A., Zorina N., Bondarenko N. Extreme genome and nervous system streamlining in the invertebrate Parasite *Intoshia variabili*. *Curr Biol*. 2020;30(7):1292-1298.e3. doi 10.1016/j.cub.2020.01.061

doi 10.18699/bgrs2024-1.1-16

COGcollator 2.0: an improved web server for analysis of distant relationships between homologous protein families

Dibrova D.V.[1, 2]*, Rykov S.Y.[3]

[1] *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

[2] *School of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

[3] *Institute of Philosophy, RAS, Moscow, Russia*

* *udavdasha@belozersky.msu.ru, dibrova.daria@yandex.ru*

Key words: clusters of orthologous groups; phylogenomic analysis; web server; domain structure; proteins

Motivation and Aim: Functional analysis of a protein sequence starts with similarity searches. A decent similarity points to the homology between protein sequences, but does not clarify if they are "the same" protein, and in what extent could the function be transferred from one homologous protein to another. Proteins usually have complex relationships between each other: homologous proteins could be either orthologs or paralogs, which originate through gene duplication and usually diverge more substantially during subsequent specification compared to orthologs. Genes could be lost during evolution, and the horizontal gene transfer which occurs frequently in prokaryotes contribute to gene gain process. These factors provide a sophisticated picture of protein phylogenetic relationships and occurrence in genomes. An approach of constructing Clusters of Orthologous Groups (COGs) and assigning proteins to them proved very fruitful from the beginning [1] and is widely applied until now [2]. We previously created a visual access tool to the data in the COG database, specifically to investigate relationships between different COGs [3]. Here, we provide its major update with an improved web interface which matches the last release of the COG database and has several important additions to our original tool.

Methods and Algorithms: We based our research on an assignment of proteins from 1309 representative prokaryotic genomes to COGs [2]. We constructed multiple sequence alignments of proteins belonging to each COG with Muscle 3.8.31 [4], and corrected multiple alignments when needed. Then we performed a search with the HMMer 3.1 [5] for each of profile HMMs on three sequence samples (Fig. 1): (1) proteins assigned to COGs in 1309 prokaryotic genomes, (2) proteins encoded in the same genomes which are not assigned to COGs, (3) eukaryotic proteins from the representative sample of 102 genomes. In multiple cases when genomes of organelles were missing from eukaryotic genome assembly, we searched for these genomes and added them manually (this was done for 36 mitochondrial genomes and 5 chloroplast genomes). Our tool shows two graphs for each COG, one with prokaryotic hits and another with eukaryotic hits of corresponding profile HMM. Each graph shows how the overall *hmmsearch* score decreases with an increase of hit rank. Both graphs are interactive: a user can obtain a protein sequence from each individual dot or multiple sequences from a range of dots, zoom graphs, browse additional information about hits, hide data which correspond to any COG in prokaryotic panel or any taxon in eukaryotic panel.

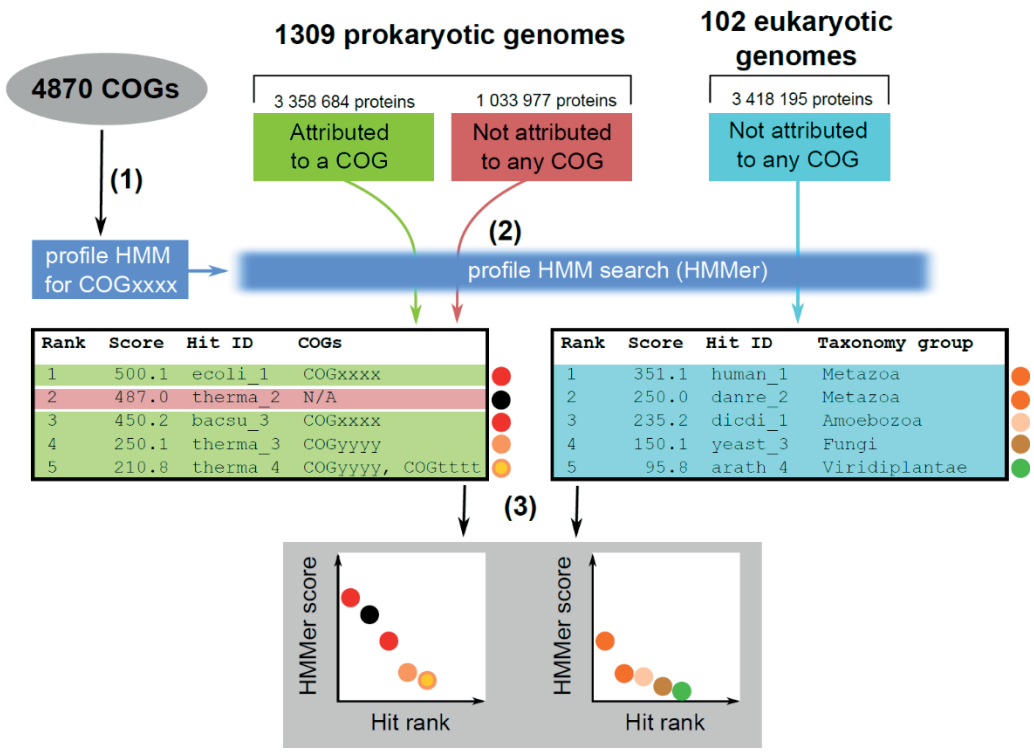


Fig. 1. Pipeline of the COGcollator 2.0. **(1)** Profile HMMs are obtained for each COG either from automatically constructed multiple alignments as described before [3], or from manually curated alignments. **(2)** Ranked hit tables for a COG are obtained for each sequence database, prokaryotic hit tables are mixed, ranked and depicted separately from the eukaryotic hit table. **(3)** Prokaryotic proteins are colored according to the COGs to which they are assigned and eukaryotic proteins are colored according to the high-level taxonomy of corresponding species

Results: We constructed profile HMM for almost all COGs from the latest COG database release (4870 out of 4877), then we calculated sensitivity of each profile and tried to improve profiles which poorly identified proteins assigned to corresponding COGs by manual inspection of alignments (126 cases it total). Detailed information on a process of each profile construction is available through the COGcollator interface. Then we performed a search with the HMMer [4] for each of profile HMMs on three sequence samples, as described above. Example of the COGcollator interface with the result for COG0055 (the catalytic β-subunit of the F_OF_1-type ATP synthase) is shown in Figure 2. It is possible to switch between two e-value thresholds for hits to be shown (1e-1 and 1e-5), thus we enabled non-strict threshold 0.1 to search for distantly related sequences and/or truncated versions of proteins. We also allowed a user to filter out COG annotations which do not intersect by protein coordinates with the hits of target COG profile HMM. This reduces number of shown COGs which are found only as fusions to the target COG.

Conclusion: The web server for COGcollator 2.0 is freely available at: http://boabio.belozersky.msu.ru/en/COGcollator. All constructed profile HMMs for 4870 COGs are united in the database and are available for download. We also provide another option to access our data: DomainAnalyser, a highly versatile sequence search tool which is based on the *hmmscan* from HMMer package and allows to visualize a domain structure of a protein set in both interactive and downloadable way.

DomainAnalyser supports our database of profile HMM for COGs as well as a widely used Pfam domain database [6]. This tool is freely available at: http://boabio.belozersky.msu.ru/en/DomainAnalyser.

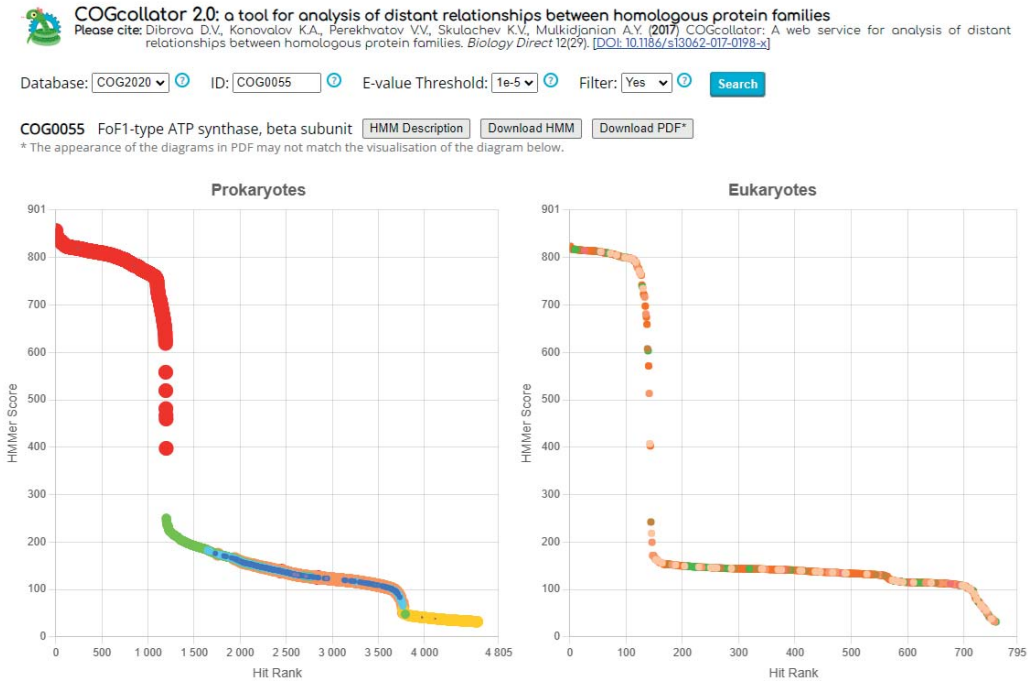


Fig. 2. Part of the COGcollator 2.0 interface: HMMer score vs. hit rank graphs for the COG0055 profile HMM at e-value threshold of 1e-5. **Left panel**, prokaryotic hits. The dots are colored according to the COG database assignment. **Right panel**, eukaryotic hits. The dots are colored according to the taxonomy of the corresponding organism

Funding: The development of original version of COGcollator and idea behind it was supported by the Dmitry Zimin "Dynasty" Foundation (D.V.D.). COGcollator 1.0 was supported by the Russian Science Foundation (RSCF), grant 14–50–00029 (D.V.D.).

References

1. Tatusov R.L., Koonin E.V., Lipman D.J. A genomic perspective on protein families. *Science.* 1997;278(5338):631-637. doi 10.1126/science.278.5338.631
2. Galperin M.Y., Wolf Y.I., Makarova K.S., Vera Alvarez R., Landsman D., Koonin E.V. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 2021;49(D1):D274-D281. doi 10.1093/nar/gkaa1018
3. Dibrova D.V., Konovalov K.A., Perekhvatov V.V., Skulachev K.V., Mulkidjanian A.Y. COGcollator: a web server for analysis of distant relationships between homologous protein families. *Biol Direct.* 2017;12(1):29. doi 10.1186/s13062-017-0198-x
4. Edgar R.C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004;5:113. doi 10.1186/1471-2105-5-113
5. Potter S.C., Luciani A., Eddy S.R., Park Y., Lopez R., Finn R.D. HMMER web server: 2018 update. *Nucleic Acids Res.* 2018;46(W1):W200-W204. doi 10.1093/nar/gky448
6. Mistry J., Chuguransky S., Williams L., Qureshi M., Salazar G.A., Sonnhammer E.L.L., Tosatto S.C.E., Paladin L., Raj S., Richardson L.J., Finn R.D., Bateman A. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021;49(D1):D412-D419. doi 10.1093/nar/gkaa913

doi 10.18699/bgrs2024-1.1-17

Absence of 5mC in thermo-tolerant yeast and telomere-to-telomere assembly by nanopore technology

Eremin A.*, Sergeev A., Malyushev D., Rodin V., Panova T., Polyakov I., Zvereva M.
Lomonosov Moscow State University, Moscow, Russia
** andreyeremin9903@gmail.com*

Key words: long-read sequencing; promethion; ogataea parapolymorpha dl-1; assembly improvement; comparative genomics; methylation

Motivation and Aim: A thermotolerant methylotrophic yeast (Ogataea parapolymorpha DL-1) has wide-ranging applications in biotechnology, from protein production to methanol metabolism studies [1, 2]. Despite its utility, gaps exist in our understanding of its genome and methylome, particularly in 5mC DNA modification and genome structure. This study seeks to resolve these gaps using long-read sequencing technology, aiming to provide a comprehensive genome assembly and an in-depth methylome analysis for 5mC. We hypothesize that this approach will reveal novel insights into the yeast's genomic architecture and regulation, which could enhance its biotechnological applications [3].
Methods and Algorithms: Nanopore sequencing technology (long-read sequencing technology of PromethION) was used to obtain T2T assembly of Ogataea parapolymorpha DL-1 genome and its methylation patterns. The process involved extracting DNA, preparing libraries, and sequencing on a PromethION device. The raw data underwent basecalling, trimming, quality checks, genome assembly and polishing. The resulting assembly was compared with previously reported draft [4] and methylation patterns of 6mA, 5mC and 5hmC were analyzed.
Results: The study resulted in a comprehensive T2T assembly of the Ogataea parapolymorpha DL-1 genome. The final assembly includes seven chromosomes and one mitochondrial genome. No plasmids were assembled from the reads. Methylation analysis revealed an absence of 5mC across the entire genome of Ogataea parapolymorpha DL-1, with only 0.01 % modifications noted, which aligns with potential sequencing errors rather than true methylation events due to 5hmC impossibly exceeding it's precursor (Table 1).

Table 1. Comparative analysis of total number modifications (5mC, 5hmC and 6mA) found with dorado. Percentage in brackets shows the part from all of the A or C in the genome

	5mC	5hmC	6mA
Ogataea parapolymorpha DL-1	123 (0.0057 %)	614 (0.0286 %)	21050 (0.9 %)
Escherichia coli	24181 (1.68 %)	1669 (0.12 %)	82863 (5.92 %)

Sequence motif TCCACCA was found in the regions ±10 pb from 6mA sites in Ogataea parapolymorpha DL-1 (Table 2). The site itself wasn't methylated suggesting that it is used by methyltransferases to modify adenosines nearby. In the validation experiment with *E. coli* known motifs GmATC and CmC(T/A)GG were found as previously reported [5].

Table 2. Modifications analysis. Motifs found with MEME in regions ±10 pb from methylated sites

	Sequence of motif	Number of motifs in regions ±10 pb from methylated sites	E-value
Ogataea parapolymorpha DL-1	TCCACCA	875	4.2e-021
Escherichia coli	GmATC	41833	1.2e-384
Escherichia coli	CmC(T/A)GG	21357	1.9e-1156

Table 3. Count of non-telomeric dT addition derived from reads aligned to the 3'end of the chromosomes

Region	A_end	T_end	perc_A	perc_T	N_total_reads
chr1:990461-990594	3	37	5.17	63.79	58
chr2:994069-994173	3	78	2.8	72.9	107
chr3:1276403-1276589	6	117	3.87	75.48	155
chr4:1299816-1299982	13	126	7.3	70.79	178
chr5:1332697-1332871	6	107	4.05	72.3	148
chr6:1517270-1517420	4	68	4.26	72.34	94
chr7:1515437-1515541	10	69	8.7	60	115

Note. A_end – number of Adenosines at the 3'end, T_end – number of Timidines respectively. perc_A – percentage of 3'end aligned reads containing dA at the end. N_total_reads – total reads, aligned to the 3'end.

Uniform distribution of 6mA modifications was observed across the chromosomes, with an average density of approximately 2 modifications per kilobase. Telomere analysis indicated that non-telomeric dT additions were present at the 3' ends of all chromosomes showing an unusual regulatory mechanism (Table 3).

The study also confirmed the presence of the previously reported [4] GGTGGCGG telomere motif at the 3' ends and identified a CCACCCCG motif at the 5' ends, which hasn't been reported before. The lengths of telomeric regions were found to be extended compared to earlier draft assembly.

Conclusion: Our comprehensive investigation into the Ogataea parapolymorpha DL-1 genome through long-read sequencing technology has advanced our understanding of this organism's genomic architecture and methylation patterns. By achieving a telomere-to-telomere assembly, we have closed previous gaps and also extended the known lengths of chromosomes, offering a complete and continuous genomic map. The absence of 5-methylcytosine across the genome suggests unique methylation dynamics and an evolutionary divergence from typical eukaryotic methylation patterns. These findings underscore the power of long-read sequencing to reveal detailed and accurate genomic features that are often obscured by short-read sequencing technologies. This work not

only enriches our genomic knowledge of Ogataea parapolymorpha DL-1 but also contributes to the broader understanding of genomic structure and stability in eukaryotes, paving the way for future studies on genomic and epigenetic diversity. Genomic assembly derived in this study as well as the sequencing reads are available at NCBI under the BioProject ID PRJNA1091144.

Funding: The study is supported by the project of the Interdisciplinary Scientific and Educational School of Moscow State University "Molecular technologies of living systems and synthetic biology" (# 23-Sh04-45).

References

1. Liebal U.W., Fabry B.A., Ravikrishnan A. et al. Genome-scale model reconstruction of the methylotrophic yeast Ogataea polymorpha. *BMC Biotechnol*. 2021;21:23. doi 10.1186/s12896-021-00675-w
2. Smekalova E.M., Malyavko A.N., Zvereva M.I. et al. Specific features of telomerase RNA from Hansenula polymorpha. *RNA*. 2013;19:1563-1574. doi 10.1261/rna.038612.113
3. Xie L., Yu W., Gao J. et al. Ogataea polymorpha as a next-generation chassis for industrial biotechnology. *Trends Biotechnol*. 2024:S0167-7799(24)00086-6. doi 10.1016/j.tibtech.2024.03.007
4. Ravin N.V., Eldarov M.A., Kadnikov V.V. et al. Genome sequence and analysis of methylotrophic yeast Hansenula polymorpha DL1. *BMC Genomics*. 2013;14:837. doi 10.1186/1471-2164-14-837
5. Kahramanoglou C., Prieto A.I., Khedkar S. et al. Genomics of DNA cytosine methylation in Escherichia coli reveals its role in stationary phase transcription. *Nat Commun*. 2012;3:886. doi 10.1038/ncomms1878

doi 10.18699/bgrs2024-1.1-18

Ultra-Conserved Non-Coding Elements in the human genome. Their possible functions and association with diseases

Fedorova L.[1], Fedorov A.[1, 2, 3]*

[1] *CRI Genetics LLC, Santa Monica, CA 90404, USA*
[2] *Program of Bioinformatics and Proteomics/Genomics, University of Toledo, Toledo, OH 43606, USA*
[3] *Department of Medicine, University of Toledo, Toledo, OH 43606, USA*
* *alexei.fedorov@utoledo.edu*

Key words: nearest neighbor; DNA conformation; meiosis; bioinformatics; ultraconserved

Thousands of prolonged sequences of human ultra-conserved non-coding elements (UCNEs) share only one common feature: peculiarities in the unique composition of their dinucleotides. Here we investigate whether the numerous weak signals emanating from these dinucleotide arrangements can be used for computational identification of UCNEs within the human genome. For this purpose, we analyzed 4272 UCNE sequences, encompassing 1,393,448 nucleotides, alongside equally sized control samples of randomly selected human genomic sequences. Our research identified nine different features of dinucleotide arrangements that enable differentiation of UCNEs from the rest of the genome. We employed these nine features, implementing three Machine Learning techniques -- Support Vector Machine, Random Forest, and Artificial Neural Networks -- to classify UCNEs, achieving an accuracy rate of 82-84%, with specific conditions allowing for over 90% accuracy. Notably, the strongest feature for UCNE identification was the frequency ratio between GpC dinucleotides and the sum of GpG and CpC dinucleotides. Additionally, we investigated the entire pool of 31,046 SNPs located within UCNEs for their representation in the ClinVar database, which catalogs human SNPs with known phenotypic effects. The presence of UCNE-associated SNPs in ClinVar aligns with the expectation of a random distribution, emphasizing the enigmatic nature of UCNE phenotypic manifestation. We propose that the key to the cryptic properties of UCNEs is hidden in their specific DNA conformations. Our hypothesis is that specific non-canonical DNA conformation of UCNEs may be integral to the homologous pairing of double-stranded DNAs during meiosis.
Funding: This research received no external funding.

References

1. Crossley E.R., Fedorova L., Mulyar O.A., Freeman R., Khuder S., Fedorov A. Computational Identification of Ultra-Conserved Elements in the Human Genome: A Hypothesis on Homologous DNA Pairing. *NAR Genom Bioinform.* 2024;6(3):lqae074. doi 10.1093/nargab/lqae074
2. Fedorova L., Crossley E.R., Mulyar O.A., Qiu S., Freeman R., Fedorov A. Profound Non-Randomness in Dinucleotide Arrangements within Ultra-Conserved Non-Coding Elements and the Human Genome. *Biology*. 2023;12:1125. doi 10.3390/biology12081125
3. Fedorova L., Mulyar O.A., Lim J., Fedorov A. Nucleotide Composition of Ultra-Conserved Elements Shows Excess of GpC and Depletion of GG and CC Dinucleotides *Genes (Basel*). 2022;13(11):2053. doi 10.3390/genes13112053

doi 10.18699/bgrs2024-1.1-19

Systematic analysis of tandem genome amplification events in *Yersinia pestis*

Konanov D.N.[1], Liubimova O.N.[2], Kovrizhnikov A.V.[3], Sonets I.V.[1], Balykova A.N.[3], Lukina-Gronskaya A.V.[1], Speranskaya A.S.[1], Eroshenko G.A.[3], Ilina E.N.[1], Govorun V.M.[1], Kutyrev V.V.[3]

[1] *Research Institute for System Biology and Medicine of Rospotrebnadzor, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Russia*

[3] *Russian Research Anti-Plague Institute "Microbe", Saratov, Russia*

* *konanovdmitriy@gmail.com*

Key words: tandem genome amplification; *Yersinia pestis*

Motivation and Aim: Tandem genome amplification is quite common in bacteria growing under stress conditions [1] while spontaneous genome amplification events are rare, not stable and generally poorly described. In most cases, tandem amplification is a result of RecA-dependent recombination which requires short DNA repeats flanking the amplified region. Plague pathogen *Yersinia pestis* is a unique organism that contains an enormous number of short repeat sequences in its genome [2] and as a result is prone to spontaneous genome rearrangements including large tandem duplication events. Now, rearrangement dynamics in *Y. pestis* is considered as one of its main adaptation mechanisms [3], and we suppose that genome amplification events also might play a role in it by affecting gene expression. In this study, we aimed to systematically consider the frequency of spontaneous genome amplification events in *Y. pestis*, identify their main drivers and analyze the gene content of the amplified regions.

Methods and Algorithms: During this study, 11 *Y. pestis* strains were sequenced using both Illumina and Oxford Nanopore platforms. A new pipeline was developed for the detection of amplified regions based on raw reads and annotation of the potential amplification drivers. In addition to our data, more than 1000 *Y. pestis* read archives were downloaded from the SRA database and processed by the new pipeline. Potential amplification drivers were identified using ICEscan, barrnap and blastn tools.

Results: This study was initiated when we observed a few coverage anomalies in the genome assemblies of 11 sequenced *Y. pestis* strains. Genomes of strains 1627, 1728, 1815, and 1818 had large chromosome regions with local coverage depth significantly higher than the median chromosome coverage depth. The common feature of these regions was flanking them by a pair of identical IS-elements that indicates that the copy number increase could be caused by either RecA-dependent site-specific recombination or "copy-and-paste" transposition [4]. Based on long reads, we inferred that each of these over-coverage events had a structure of tandem genome amplification. Families of IS-elements triggering the recombination in our genomes were annotated as IS3 and IS200/IS605. Since these IS-families are most represented in *Y. pestis* genomes, we could not conclude if there are the preferred amplification drivers, so we decided to expand the search space and downloaded more than one hundred raw read archives from SRA marked as *Y. pestis* whole-genome sequencing data. Firstly, we processed about 100 randomly selected read archives manually and found that over-coverage events appear in at least half of sequenced *Y. pestis* isolates. In rare cases, the amplified regions

were flanked by rRNA gene clusters, which means that these tandem amplifications were more likely the results of RecA-dependent recombination than transposition. Next, we processed all collected data using a new developed pipeline that included the detection of amplified regions and identification of potential amplification drivers.

We found that there are no preferred types of IS-elements or any other repeated sequences triggering the amplification. In most cases, amplified regions were reproducible only between samples belonging to one bioprojet but a few regions were reproducible between different bioprojects. Thus, the 46 kb amplified region that we detected in our 1627 strain was found in *Y. pestis* PBM19. Interestingly, the genome assembly available in RefSeq for this strain does not contain the tandem repeat despite usage of long reads.

After the amplified regions were identified and annotated, we performed the gene function enrichment to check if there are metabolic pathways that are preferably over-covered. We did not find significant over-representation of individual pathways in the analyzed regions so the amplified regions seem to be randomly located on the chromosome and depend only on the presence of repeat sequences. The length and the copy number increase rate of the amplified regions were diverse: the longest identified fragments had the length higher than 1 Mb, and the highest number of tandem repeats was more than 20x. Expectedly, higher copy number was mostly observed for shorter regions that is likely explained by change in fitness cost [5]. Independently, the regions with significantly lowered coverage depth were considered. The only reproducible under-covered regions were *pgm*, and *Y. pestis* filamentous phage [6]. According to the published results, *pgm* is known to be occasionally eliminated under laboratory conditions in curtain *Y. pestis* strains [7], while the filamentous phage is stably integrated into the chromosome only in 1.ORI strains and almost exclusively extrachromosomal in other *Y. pestis* strains [8].

We did not find clear evidence of stable tandem amplification in bacteria different from *Y. pestis* since all described amplification events were associated with stress conditions. Even species close to *Y. pestis* such as *Y. pseudotuberculosis* and *Y. enterocolitica* do not have in their genomes such number of short repeats and the only described case of tandem duplication in them is an amplification of 28 kb fragment containing *blaA* gene in *Y. enterocolitica* appeared under antibiotic stress [9]. Thus, it makes it more intriguing to next investigate the mechanisms and the adaptive role of such copy number dynamics in the plague pathogen.

Funding: This work did not receive any financial support from commercial or non-commercial organizations.

References

1. Andersson D.I., Hughes D. Gene amplification and adaptive evolution in bacteria. *Annu Rev Gen*. 2009;43:167-195
2. Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G. et al. Genome sequence of Yersinia pestis, the causative agent of plague. *Nature*. 2001;413(6855):523-527
3. Bochkareva O.O., Dranenko N.O., Ocheredko E.S., Kanevsky G.M., Lozinsky Y.N. et al. Genome rearrangements and phylogeny reconstruction in Yersinia pestis. *Peer J*. 2018;6:e4545
4. Reams A.B., Roth J.R. Mechanisms of gene duplication and amplification. *Cold Spring Harbor Perspect Biol*. 2015;7(2):a016592
5. Ravatn R., Studer S., Springael D., Zehnder A.J., van der Meer J.R. Chromosomal integration, tandem amplification, and deamplification in Pseudomonas putida F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from Pseudomonas sp. strain B13. *J Bacteriol*. 1998;180(17):4360-4369

6. Derbise A., Chenal-Francisque V., Pouillot F., Fayolle C., Prévost M.C. et al. A horizontally acquired filamentous phage contributes to the pathogenicity of the plague bacillus. *Mol Microbiol*. 2007;63(4):1145-1157
7. Leal-Balbino T.C., Leal N.C., Nascimento M.G., Oliveira M.B. et al. The pgm locus and pigmentation phenotype in Yersinia pestis. *Gen Mol Biol*. 2006;29:126-131
8. Chouikha I., Charrier L., Filali S. et al. Insights into the infective properties of YpfΦ, the *Yersinia pestis* filamentous phage. *Virology*. 2010;407(1):43-52
9. Seoane A., Sánchez E. et al. Tandem amplification of a 28-kilobase region from the Yersinia enterocolitica chromosome containing the blaA gene. *Antimicrob Agents Chemother*. 2003;47(2):682-688

doi 10.18699/bgrs2024-1.1-20

Approaches for the coronavirus genomes annotation

Korneenko E.V.[1, 2]*, Chudinov I.K.[1, 3], Butenko I.O.[1], Speranskaya A.S.[1]

[1] *Scientific Research Institute for Systems Biology and Medicine, Federal Service on Consumer Rights Protection and Human Well-Being Surveillance, Moscow, Russia*

[2] *Saint Petersburg Pasteur Institute, Federal Service on Consumer Rights Protection and Human Well-Being Surveillance, St. Petersburg, Russia*

[3] *Moscow Institute of Physics and Technology (National Research University), Moscow, Russia*

* *lenakorneenko0@gmail.com*

Key words: viruses; coronaviruses; genome annotation

Motivation and Aim: The discovery of novel viruses especially coronaviruses (CoVS) has great importance in public health [1]. After the genome assembly each sequence should be annotated. Accurate gene prediction is a challenge that persists for decoding a newly sequenced genome [2]. To describe and visualize features of newly obtained viral genomes we applied the three step approach which included similarity-based prediction of genes/cds, protein domains prediction, output verification and manual correction.
Methods and Algorithms: Similarity-based prediction of genes/cds for assembled coronaviruses was performed using several programs such as Geneious Prime (https://www.geneious.com), VIGOR4 (a part of BV-BRC web service) [3] and Prokka, protein domains were predicted using the InterProScan tool against InterPro database [4]. Two alphacoronaviruses (Alpha-CoVs) belonging to *Pedacovirus* subgenus found in bats were annotated using Geneious Prime and Prokka [5]. Two betacoronaviruses (Beta-CoVs) belonging to the *Merbecovirus* subgenus found in hedgehogs were annotated using BV-BRC and Prokka. The output of annotators was manually checked, compared to the closest reference annotation using NCBI blast n, blast x, blast p and corrected.
Results: We performed gene/cds and protein domain prediction for 4 newly discovered complete genomes of coronaviruses: 2 Alpha-CoVs found in bats and 2 Beta-CoVs found in hedgehogs. We could predict 7 of 7 expected cds in *Pedacovirus* subgenus of Alpha-CoVs with Geneious prime, while Prokka found 8 cds. BV-BRC showed the presence of 10 of 11 mandatory cds for Beta-CoVs related to the *Merbecovirus* subgenus and Prokka showed the presence of 7 of 11 expected cds. Protein domain prediction using InterProScan successfully predicted typical Alpha-CoV domains in non-structural proteins. Protein domain annotation revealed that two Alpha-CoVs both contained large insertions of 130+ a.a. in the Nsp3 sequence. This region was labeled as “protein domain with unknown function”.
Conclusion: Annotation programs suitable for virus genomes (Prokka, BV-BRC and Geneious Prime) predicted most of expected cds in newly assembled CoVs as well as InterProScan was successful in annotation of protein domains in these sequences, but accurate gene and protein domain prediction required detailed knowledge of *Coronaviridae* genome structure, manual checking and correction. Thus, development of specialized algorithms for coronaviral genomes’ annotation that account for all known phenomena is a relevant task.

Funding: The study is supported by the state project “Development of algorithms to identify novel, unique DNA or RNA sequences in metagenomes and their phenotypic characterization in vitro” (No. 12203090069-4).

References

1. van Boheemen S., de Graaf M., Lauber C. et al. Genomic Characterization of a Newly Discovered Coronavirus Associated with Acute Respiratory Distress Syndrome in Humans. *mBio*. 2012;3(6):e00473-12. doi 10.1128/mBio.00473-12
2. Mills R., Rozanov M., Lomsadze A. et al. Improving Gene Annotation of Complete Viral Genomes. *Nucleic Acids Res.* 2003;31(23):7041-7055. doi 10.1093/nar/gkg878
3. Olson R.D., Assaf R., Brettin T. et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): A Resource Combining PATRIC, IRD and ViPR. *Nucleic Acids Res.* 2023;51(D1):D678-D689. doi 10.1093/nar/gkac1003
4. Paysan-Lafosse T., Blum M., Chuguransky S. et al. InterPro in 2022. *Nucleic Acids Res*. 2023;51(D1):D418-D427. doi 10.1093/nar/gkac993
5. Seemann T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics*. 2014;30(14):2068-2069. doi 10.1093/bioinformatics/btu153

doi 10.18699/bgrs2024-1.1-21

Bacterial genomes are internally formatted

Korotkov E.V.*

Institute of Bioengineering, Research Center of Biotechnology, RAS, Moscow, Russia

** bioinf@yandex.ru*

Key words: genome; dispersed repeats; bacteria

Motivation and Aim: The goal was to try to find dispersed repeats in the bacterial genome. It is known that the bacterial genome is packaged into a nucleoid through the interaction of DNA with certain proteins. Due to this packaging, it can be assumed that weakly similar families of dispersed repeats may exist in the bacterial genome, which could not be detected by previously developed computer methods. Therefore, we decided to look for dispersed repeats in bacterial genomes in the range $1.0 \leq x \leq 1.7$, where x is the average number of mutations per nucleotide between any two repeats from a family.

Methods and Algorithms: We have developed a de novo method for the identification of dispersed repeats based on the use of random position-weight matrices (PWMs) and an iterative procedure (IP) [1]. The created algorithm (IP method) allows detection of dispersed repeats for which the average number of substitutions between any two repeats per nucleotide (x) is less than or equal to 1.7. IP method made it possible to detect families of dispersed repeats in bacterial genomes which have not been previously reported. The IP method uses positional weight matrices, dynamic programming and an iterative procedure. Z, which is approximately an argument for a normal distribution, will be used as an estimate of statistical significance.

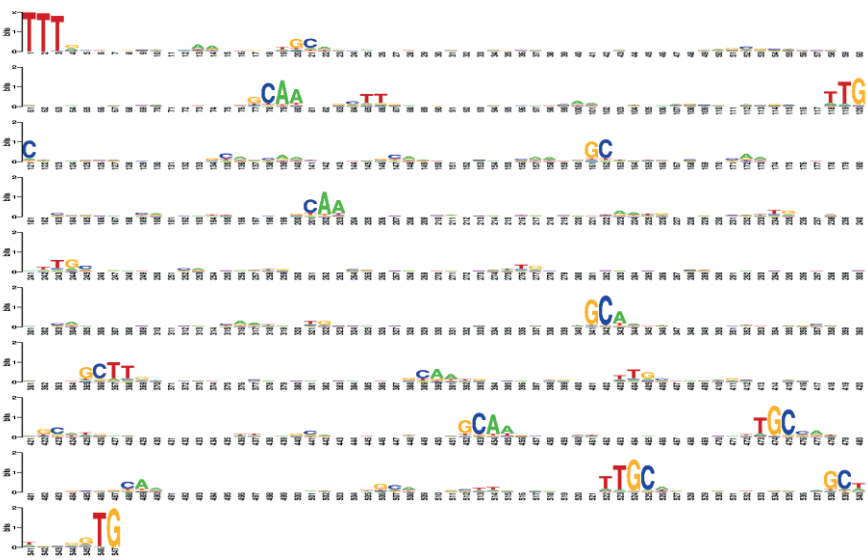


Fig. 1. Consensus sequence for the dispersed repeat family from the *Clorobium chlorochromatii* genome constructed using Weblogo [2]

Results: We applied this method to find dispersed repeats in the genomes of *E. coli* and 42 bacteria from other bacterial phyla and could identify repeat families comprising from 1.0×10^3 to 1.4×10^4, depending on the species. Each bacteria contains only one family of dispersed repeats with lengths from 440 to 580 bases. The repeats found occupy from 30 to 60 % of the bacterial genome, and more than 90 % of the repeats occur in the coding sequences of the genome. They are superimposed on the coding region as a motif. Such extensive repeat family could not be detected in the analyzed genomes by using the RED, RECON, or Repeat_masker programs but only by the IP method, which could find de novo repeat families with $x \leq 1.7$, whereas all other programs could do it with $x \leq 1.0$.

In Figure 1 we show the consensus sequence that was constructed using the WebLogo program [2] for a family of repeats from the Clorobium chlorochromatii genome. The figure shows that dispersed repeats of this family diverged greatly from each other. However, one can notice a fairly large number of short regions of 2-5 bases in length, which are almost identical in all repeats. The existence of such repeats shows that bacterial genomes are marked up like a computer hard drive. This markup could be involved in the creation of the liquid crystal structure within bacterial DNA through interactions between repeats within a family and some proteins [3].

Funding: The study is supported by RNF grant 24-24-00031.

References

1. Korotkov E., Suvorova Y., Kostenko D., Korotkova M. Search for dispersed repeats in bacterial genomes using an iterative procedure. *Int J Mol Sci*. 2023;24:10964. doi 10.3390/IJMS241310964/S1
2. Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. WebLogo: a sequence logo generator. *Genome Res*. 2004;14:1188-1190. doi 10.1101/GR.849004
3. Yevdokimov Y.M., Salyanov V.I., Skuridin S.G. From liquid crystals to DNA nanoconstructions. *Mol Biol*. 2009;43:284-300. doi 10.1134/S0026893309020113/METRICS

doi 10.18699/bgrs2024-1.1-22

The synergistic effect of p53 and PARP-1 on the nucleosome

Koshkina D.O.[1]*, Maluchenko N.V.[1], Novichkova A.M.[1], Feofanov A.V.[1, 2], Studitsky V.M.[1, 3]

[1] *Lomonosov Moscow State University, Moscow, Russia*

[2] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

[3] *Fox Chase Cancer Centre, Philadelphia, Pennsylvania, USA*

* *koshkina.daria.2015@post.bio.msu.ru*

Key words: chromatin; p53; PARP-1; nucleosome; epigenetics

Motivation and Aim: Both poly (ADP-ribose) polymerase 1 (PARP-1) and p53 have been extensively studied due to their regulatory role in numerous pivotal biologic pathways. It is assumed that they are responsible for DNA damage response, differentiation, metabolism, and aging. In addition to their common functions, they interact with each other. Nevertheless, the details of PARP-1 and p53 chromatin- reorganizing activity are still uneven. The aim is to investigate the combined effect of proteins (PARP-1 and p53) on the nucleosome.

Methods and Algorithms: Fluorescently labeled mononucleosomes were assembled on a DNA template containing the strong nucleosome positioning 603 Widom flanked by two 20-bp linker DNA fragments. The templates were obtained by polymerase chain reaction using the fluorescently labeled oligonucleotides. The fluorescent pair Cy3/Cy5 was set at 10/10 bp (from the beginning of the 603 sequence) at the linker region of the nucleosome. Formation of complexes between proteins (PARP-1/p53DBD) and nucleosomes were proceeded using a potassium-containing buffer with physiologically accurate concentration of K+ (150 mM). Obtained complexes were studied by the electrophoretic mobility shift assay using native PAGE.

Results: In the present study, we conducted a series of experiments with the following structure.

During the first series of experiments, p53 was added to pre-incubated nucleosome-PARP-1 (50 nM) complexes at increasing concentrations of p53 (25-200nM). During the second series of experiments, PARP-1 (50 nM) was added to pre-incubated nucleosome-p53 (25-200nM) complexes. In both cases, a set of complexes of different stoichiometry was obtained. P53 exhibited higher activity when interacting with the nucleosome-PARP-1 complex, and PARP-1 when the reverse method of setting up the experiment was used.

Conclusion: Thus, we can assume a possible synergistic interaction of proteins during the formation of complexes with the nucleosome, as well as an independent one with displacement of one of the agents.

Funding: The study is supported. This study was supported by the Russian Science Foundation (No. 19-74-30003).

doi 10.18699/bgrs2024-1.1-23

Genome analysis of novel strains of *Cydia pomonella* granulovirus from Russia

Lakhova T.[1, 2, 3]*, Klimenko A.[1, 2, 3], Tsygichko A.[4], Ismailov V.[4], Vasiliev G.[1, 2], Asaturova A.[4], Lashin S.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

[4] *Federal State Budgetary Scientific Institution "Federal Research Center of Biological Plant Protection", Krasnodar, Russia*

** tlakhova@bionet.nsc.ru*

Key words: *Cydia pomonella* granulovirus; codling moth; genome sequencing; comparative genomic analysis

Motivation and Aim: *Cydia pomonella* is a pest of fruit trees such as apple and pear, as it damages their fruits and thus causes damage to agriculture. *Cydia pomonella* granulovirus (CpGV) is a natural pathogen for *C. pomonella* that is used as a biocontrol agent of the insect population. One of the earliest isolates of CpGV isolated and widely used to control *C. pomonella* populations is the Mexican isolate. However, it has recently been observed that some natural moth populations may exhibit resistance to formulations based on well-known CpGV isolates. Thus, the study of novel CpGV isolates and comparison with known isolates is necessary for identifying promising isolates that could form the basis for highly effective bioinsecticides in the future. In our study, we investigated 16 new strains obtained from infested *C. pomonella* caterpillars collected in Russia (Krasnodar Krai and Rostov Oblast) and 2 strains from Kazakhstan.

Methods and Algorithms: Quality control of raw data was carried out by FastQC and fastp. We used the NC_002816.1 genome from the NCBI RefSeq database as a reference virus genome. BWA is used for mapping our data to the reference genome. Draft assemblies are performed by two assemblers: Spades for *de novo* assembly and MinYS for reference-guided assembly. To reduce possible assembly errors, the Pilon tool is applied to each assembly. To produce a more complete assembly, genome polishing was performed with the Gfinisher tool based on the draft assemblies. Quast is applied after each assembly step to control the quality of the assembly. Genome annotation is performed using the Prokka tool. We compared our genomes to CpGV genomes from NCBI Virus that have complete assemblies using average nucleotide identity (ANI) with FastANI tool. We also used progressiveMauve to examine multiple alignments of our genomes relative to NC_002816.1 to identify gene rearrangement or loss, as well as regions of low similarity. Comparison of gene content with two reference genomes: NC_002816.1 and KM217575 was performed using GeneAPI. In addition, we used CNVpytor to confirm the absence of genes listed in GenAPI that may have resulted from deletions.

Results: We obtained 18 assemblies consisting of four contigs on average. FastANI demonstrated high values (ANI>99 %) against 14 genomes from NCBI Virus. Multiple alignments of the genomes showed large regions with low sequence similarity in the following genomes: BZR GV 2, BZR GV 5, BZR GV 12, BZR GV L-2, BZR GV L-5

and BZR GV L-7. We consider the region to be large if its size is higher than 300 bp in length. GeneAPI showed that some genes are missing. The orf6 and IFEMGEHL_00128 genes were absent in the most isolates. Gene absence was defined if the level of gene sequence similarity was <90 % and coverage <50 %. CNV analysis confirmed deletions for the genes identified as absent in GenAPI and missing genes in progressiveMauve for BZR GV L-4 and BZR GV L-6 as true deletions. We also detected SNPs when we analyzed genes and their products involved in the virosis infectious process.

Conclusion: Comparative genomic analysis of the 18 genomes of the new strains showed high similarity to the reference isolates, but they also have several differences that probably affect the virulence of the strains or the mechanism of interaction with the codling moth. These differences require further studies to understand their impact.

Funding: The study is supported by Russian Science Foundation (No. 23-16-00260).

doi 10.18699/bgrs2024-1.1-24

Does G-Quadruplexes formation elevate the mutation density in its neighborhoods in the human genome?

Panova V.[1]*, Alexeevski A.[2, 3], Zvereva M.[4]

[1] *Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

[2] *Belozersky Institute of Moscow State University, Moscow, Russia*

[3] *Department of Mathematics, Scientific Research Institute for System Studies, Russian Academy of Sciences*

[4] *Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia*

** nooroka@fbb.msu.ru*

Key words: G-quadruplex; mutation density; mutations; cancer; pattern

Motivation and Aim: The origin of cancerous processes stems from a diverse range of mutations found in both somatic cells and germ lines. It is well known that promoter mutations in oncogenes play a significant role in the development and progression of cancer. Three G-quadruplexes (G4s) are formed *in vitro* from the *hTERT* gene promoter in a 96-nt DNA sample, even in the presence of single or double specific mutations in G-quadruplex (G4). G4 binds to the MutL component of the MMR repair system. This reduces the efficiency of repair of incorrectly paired bases compared to dsDNA [1].

The study aims to test the hypothesis that the formation of G4 leads to an increase in the frequency of mutations in the neighborhoods of G4. We have already shown this fact for G4 loop-forming sequences for *Primates* in the set of mammal *TERT* promoters in [2]. Furthermore, it has been demonstrated that in a subset of individuals with multiple myeloma, somatic mutations in genomic areas predicted to create G4 structures are more common in tumor plasma cells. Patients with G4 strong context enrichment in their tumors share certain germline SNPs [3].

Methods and Algorithms: To select experimentally observed G-quadruplexes we used data from [4]. These data are presented as coordinates of 428624 and 1285463 peaks of DNA polymerase (Observed G-Quadruplexes, OQs) stalling in the whole human genome (assembly GRCh37) and have an identifier GSM3003539 and GSM3003540 in GEO, respectively [5]. Samples from GSM3003539 contain K+, which stabilizes G4 structures, and samples from GSM3003540 additionally contain pyridostatin (PDS) added to K+. PDS is a ligand which specifically stabilizes G4-structures [6].

The method from [4] and [6] is developed for G4 detection because G4 formation is known to stall polymerase. Nevertheless, the method is not free of overpredictions. For example, we have found 3387 OQs with 0 % GC content in GSM3003539 and 9840 OQs in GSM3003540. We searched all peak sequences for the presence of canonical G4 patterns $G_{3+}\,L_{1\text{-}7}\,G_{3+}\,L_{1\text{-}7}\,G_{3+}\,L_{1\text{-}7}\,G_{3+}$. To each found G4 sequence we added 10 nt flanks on 5’ and 3’ ends and considered them G4 neighborhoods. As a control set, we used inter-quadruplex sections in the experimental data. The GC composition was calculated in the subsequent sections equal to the mean G-quadruplex length in each chromosome. Sections with GC-composition ≥ 50 % were considered. The number of considered quadruplexes for assembly GRCh37 was 186490 for GSM3003539 and 325172 for GSM3003540. We didn’t perform calculations on the assembly GRCh38, because the

liftover tool (https://genome.sph.umich.edu/wiki/LiftOver) sometimes translates initial coordinates from GRCh37 to CRGh38 improperly.
Mutations from dbSNP from GRCh37.p13 and COSMIC non-coding variants (v98) were taken into account. Mann-Whitney criteria was used to confirm the validity of the results.
Results: The Snakemake pipeline was developed for automatically processing the experimental data. It takes as an input the experimental and the whole-genome data and outputs the file with mutational density for each chromosome for G4-quadruplexes and control sets, coordinates of G4-quadruplexes and control sequences.
In G4-quadruplexes the mutation density by dbSNP is statistically significantly higher for GRCh37 (Mann – Whitney test gives a p-value, which is significantly lower than 0.05 for the both experimental data sets)
COSMIC data. which contain somatic mutations in cancer are processed now. An article is currently being written on the results obtained.
Conclusion: We found that the formation of G4 leads to an increase in the frequency of dbSNP mutations in the neighborhoods of G4. It is necessary to check whether the accumulation of somatic mutations represented in COSMIC is related to another process.
The link to the Snakemake pipeline: https://github.com/nooroka/mutation_density_pipeline_v2
Funding: This study was supported by the Russian Science Foundation (project no. 21-14-00161) and by the state budget for scientific research at Lomonosov Moscow State University.

References

1. Pavlova A.V., Savitskaya V.Y., Dolinnaya N.G. et al. G-Quadruplex Formed by the Promoter Region of the hTERT Gene: Structure-Driven Effects on DNA Mismatch Repair Functions. *Biomedicines*. 2022;10(8):1871. doi 10.3390/biomedicines10081871
2. Panova V.V., Dolinnaya N.G., Novoselov K.A. et al. Conserved G-Quadruplex-Forming Sequences in Mammalian *TERT* Promoters and Their Effect on Mutation Frequency. *Life (Basel)*. 2023;13(7):1478. doi 10.3390/life13071478
3. Zhuk A.S., Stepchenkova E.I., Zotova I.V. et al. G-Quadruplex Forming DNA Sequence Context Is Enriched around Points of Somatic Mutations in a Subset of Multiple Myeloma Patients. *Int J Mol Sci*. 2024;25(10):5269. doi 10.3390/ijms25105269
4. Marsico G., Chambers V.S., Sahakyan A.B. et al. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res*. 2019;47(8):3862-3874. doi 10.1093/nar/gkz179
5. Clough E., Barrett T. The Gene Expression Omnibus Database. *Methods Mol Biol*. 2016;1418:93-110. doi 10.1007/978-1-4939-3578-9_5
6. Chambers V.S., Marsico G., Boutell J.M. et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol*. 2015;33(8):877-881. doi 10.1038/nbt.3295

doi 10.18699/bgrs2024-1.1-25

Investigation of large tandem repeats in the genome of Grass frog (*Rana temporaria*)

Popova M.[1, 2]*, Komissarov A.[3], Podgornaya O.[2], Travina A.[2]

[1] *Center for Molecular and Cellular Biology, Skolkovo Institute of Science and Technology, Moscow, Russia*

[2] *Institute of Cytology, RAS, St. Petersburg, Russia*

[3] *Applied Genomics Laboratory, SCAMT Institute, ITMO University, St. Petersburg, Russia*

* *marinaalexpopova@yandex.ru*

Key words: tandem repeats; *Rana temporaria*; repetitive sequences

Motivation and Aim: Non-coding repetitive DNA sequences, particularly transposable elements (TEs) and tandem repeats (TRs), form significant parts of eukaryotic genomes. TRs are crucial for chromosomal functions such as kinetochore formation and genome organization. Despite their importance, TRs remain understudied, especially in non-mammalian species. Amphibians, with their large and variable genome sizes, serve as excellent models for studying TRs. The grass frog, *Rana temporaria*, occurs in Europe, including Russia, and is an ideal subject due to its sequenced genome and potential for experimental validation of computational predictions. This study aims to identify and map TRs in the *Rana temporaria* genome using bioinformatics and experimental approaches, including *in situ* hybridization, to enhance our understanding of the molecular mechanisms governing development and evolution in amphibians and contribute to the broader field of genomic architecture in vertebrates.

Methods and Algorithms: Publicly available Illumina reads from the *Rana temporaria* low-coverage genome sequence project (PRJNA294436) and genome assembly (aRanTem1.1 version from NCBI, GCF_905171775.1) were utilized. K-mers (k=23) were extracted from the raw reads after removing Illumina adapters, sorted by frequency, and assembled into circular paths using a de Bruijn graph approach with the extracTR tool. The most frequent k-mers in each path were identified and used to create labeled oligonucleotide probes for fluorescence *in situ* hybridization (FISH). To confirm the extracTR results, TAREAN was also employed on a one-million-read subset. Tandem Repeat Finder (TRF) was used to identify TRs in the genome assembly, focusing on those over 10,000 bp and excluding those within protein-coding genes. Monomers of identified TRs were converted into dimer form for further analysis. A blastn search was used to cluster repeats into families, and their similarity to transposable elements was assessed using the DFAM database. Results were visualized with Python's plotly library.

Results: Using extracTR, we identified six abundant TRs from reads, comparing these with TAREAN results. Two k-mer families detected by extracTR were not identified by TAREAN, and the reads per million estimate for one specific family (Rtem494A, also known as s1a) was lower in extracTR than in TAREAN, likely due to TAREAN's input data limitations. Six probes targeting the most abundant TRs in raw reads were used for *in situ* hybridization on chromosomes, with probes derived from the most conserved monomer sequences, except for TR Rtem494a, for which the additional PCR-amplified probe also was used due to its longer monomers. The *R. temporaria* karyotype, with 26 chromosomes, showed intense hybridization signals mainly at the centromeres of

large and some small chromosomes. Additional signals were noted on chromosome arms, particularly with Rtem47a. Rtem219a probes, the latter correlating with the non-transcribed spacer of 5S rDNA and avoiding the nucleolus organizer region on chromosome 10, displays intense signals in the centromeric regions of 5 big and 2 small chromosomes. Signal from centromere of chromosome 7 is probably to coincide to the only 5S rDNA loci, found a in the early studies [1].

Using the aRanTem1.1 reference genome, 768 tandem repeat arrays exceeding 10,000 bp were identified, grouped into 76 families or classified as singletons, with many located on unplaced or unlocalized scaffolds. The identified TR families varied in monomer lengths (4–1930 bp) and GC content (27.6–65.95 %). Four out of the six most abundant TRs in raw reads were also prominent in the genome assembly. Dfam database searches revealed no significant similarity between TR family consensus sequences and transposable elements, though some TRs showed likeness to tRNA and snRNA genes. Chromosome specificity varied among TR families, with several located on multiple or unplaced scaffolds. Despite the low representation of TRs in the reference genome (0.32 % for TRs over 10 kb), their distribution along chromosome arms and enrichment in pericentromeric and subtelomeric regions were mapped, revealing some discrepancies between *in situ* and *in silico* data, highlighting challenges in repetitive DNA assembly.

Two noteworthy findings were: (1) TRs of the Rtem219A family consist of 5S rRNA sequences and a non-transcribed spacer, and (2) probe Rtem47A, derived from raw reads, displayed strong FISH signals but was found in only a few copies across three specific TR arrays in the assembly, suggesting high repetitiveness across the genome. We hypothesize that probe Rtem47A represents a dispersed repeat containing an inner tandem repeat.

Conclusion: Repetitive DNA sequences, including TRs, are challenging to assemble and annotate due to their complex structures and prevalence in centromeric, pericentromeric, and subtelomeric regions, which are typically underrepresented in genome assemblies. All TRs found with TRF constitute about 10.23 % of the *R. temporaria* genome, with many arrays located on contigs with unresolved gaps. Assembly-free methods like TAREAN detect abundant TRs but struggle with low-copy and short dispersed TRs. Using a more extensive dataset with the extracTR tool allowed us to identify TRs that TAREAN missed. FISH successfully detected all six targeted TRs on *R. temporaria* chromosomes, validating the effectiveness of these bioinformatics approaches.

The Rtem219a-probe hybridization in the centromeric regions of several chromosomes indicates the evolutionary dynamism of rDNA, with potential spread across the genome creating new loci, non-functional pseudogenes, and possibly satellite DNA (TRs), but further detailed studies are required to clarify the nature and functionality of these 5S rRNA gene variants and pseudogenes. Such TRs, arising from 5S rDNA, have already been described for some species of plants, fish and amphibians. This underscores the complexity and evolutionary significance of rDNA organization in the genome.

Funding: The study is supported by a grant from the Russian Science Foundation (No. 24-24-00480).

References

1. Vitelli L., Batistoni R., Andronico F., Nardi I., Barsacchi-Pilone G. Chromosomal localization of 18S+28S and 5S Ribosomal RNA genes in evolutionarily diverse anuran amphibians. *Chromosoma*. 1982;84(4):475-491. doi 10.1007/BF00292849

doi 10.18699/bgrs2024-1.1-26

Deciphering the structural organization of sex chromosomes in Siberian silk moth, *Dendrolimus sibiricus*

Shipova A.A.[1], Ershov N.I.[1], Yakimova M.E.[2, 3], Martemyanov V.V.[2]*

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia*

[3] *Department of Information Biology, Novosibirsk State University, Novosibirsk, Russia*

* *martemyanov79@yahoo.com*

Key words: sex determination; whole-genome sequencing; chromosome evolution; W chromosome

Motivation and Aim: The W chromosome of *Lepidoptera* is typically saturated by genomic repeats, gene-poor and completely heterochromatinized. These properties as well as haploid state of sex chromosomes in contrast to diploid state of the rest genome significantly hamper the accurate assembly of W chromosome from heterogametic individuals. To avoid these problem, many genome sequencing projects prefer to study homogametic individuals. We explored the possibility to circumvent these problems at low cost using an approach of contig classification and targeted reassembly of available long-read genomic data obtained from female *D. sibiricus*.

Methods and Algorithms: Two short-read genomic libraries of female and male *D. sibiricus* were sequenced at low depth and mapped on the primary assembly of *D. sibiricus* female genome. The ratios of coverage by these libraries were compared to the theoretically expected values according to the ZW/ZZ sex determination scheme and used to classify the contigs on W, Z and autosomal. From the raw HiFi data, the portion mapped on W-derived contigs was extracted and assembled using the hifiasm with parameters configured for haploid assembly.

RepeatMasker and RepeatModeler software were used to predict and identify genomic repeats in the assembly and to perform Kimura 2-parameter divergence analysis. The analysis of syntenic regions between the Z and W chromosomes included their pairwise alignment by minimap2 and aggregation into syntenic blocks by Asynt R-package. Non-partitioned phylogenetic analysis of W-associated genes was performed using IQ-TREE software with 10,000 bootstrap iterations.

Results: The primary diploid assembly of *D. sibiricus* contained a significant portion of fragments that were represented in one haplovariant but not linked to the Z chromosome, i.e. potentially related to the W chromosome. Using two additional genomic short-read libraries of female and male *D. sibiricus*, we built a classifier to partition all contigs into autosomes, the Z chromosome, and the W chromosome according to the expected ZW/ZZ sex determination scheme.

The initially used hifiasm assembler is designed to assemble diploid genomes, thus it may misassemble unpaired sex chromosomes. We performed a targeted reassembly of the W chromosome, limited to only reads mapped to W-classified unitigs. Nevertheless, the assembly retained some fragmentation, accounting 10 contigs with a total length of 22.46 Mbp. Unresolved assembly fragmentation of the W chromosome is explained by its exceptional saturation with genomic repeats together with the fact that the W chromosome contained several large and extremely homologous duplications.

Accounting them, the actual length of W is expected to be 27.4 Mbp, while we kept only one copy of each in the final assembly.

Interestingly, we discovered two large fragments of the symbiotic *Wolbachia pipientis* genome as part of the W chromosome. The inserted fragments both are characterized by relative preservation of the coding ORFs, assuming that these insertions occurred quite recently.

The W chromosome sequences are also distinguished by a clear expansion of LTR transposons. The proportion of these repeats was more than 3 times higher compared to the whole genome. The total proportion of repeats in the W chromosome was about 80 %. The analysis of interspersed repeat landscape based on Kimura 2D distances revealed significant differences between the W chromosome and the rest nuclear genome, suggesting that the evolutionary age of the large-scale expansion of the LTR transposons along the W chromosome is very recent. Moreover, a series of repeats were strongly overrepresented or even specific to the W chromosome. Among them was dispersed telomere-associated non-LTR transposon accounting for almost 3 % of its sequences.

Analysis of synteny between Z and W chromosomes revealed significant regions of their collinearity at both distal subtelomeric regions. Almost all protein-coding genes predicted on the W chromosome were concentrated in these regions of synteny, having a close paralog on Z. A comparison of the Z and W paralogues indicated that most W chromosome genes are pseudogenes with a damaged ORF. The phylogenetic analysis of the preserved genes confirmed that at least part of the W chromosome was formed from a copy of Z of *D. sibiricus*.

Conclusion: The abundance of young repeats and young insertions of *Wolbachia* genomic fragments indicate the recent formation of the W chromosome, which echoes the theory of multiple non-canonical emergence of W chromosomes in the evolution of *Lepidoptera*, including its formation from B chromosomes [1]. At the same time, the W exhibits significant sequence similarity to the Z, which confirms its evolutionary origin from the Z chromosome.

References

1. Fraïsse C., Picard M.A.L., Vicoso B. The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. *Nat Commun*. 2017;8(1):1486. doi 10.1038/s41467-017-01663-5

doi 10.18699/bgrs2024-1.1-27

Identification of small RNAs stably associated with chromosome in the metaphase of 3T3 cell line

Ye B.[1, 2†], Meng Y.[3†], Zhang L.[2†], Xinhui Li[2]*, Hua Li[1, 2]**

[1] *Jiangsu Province Engineering Research Center of Molecular Target Therapy and Companion Diagnostics in Oncology, Suzhou Vocational Health College, Suzhou, China*

[2] *State Key Laboratory of Systems Medicine for Cancer, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China*

[3] *Shanghai Key Laboratory of Embryo Original Diseases, The International Peace Maternity and Child Health Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China*

†These authors contributed equally to this work.

** xhli@sjtu.edu.cn; ** kaikaixinxin@sjtu.edu.cn*

Key words: small RNAs; mouse 3T3 Cells; mitotic chromosomes; RNA sequencing

Motivation and Aim: During the different stages of mitosis, various types of noncoding RNAs (ncRNAs) that are associated with chromatin must maintain strict proportions and structures. These ncRNAs form an RNA infrastructure that is involved not only in processing other RNAs, such as mRNAs, tRNAs, and rRNAs, but also in gene regulation by targeting mRNAs and chromatin. Previous studies have revealed the lncRNA composition of mitotic chromosomes, suggesting the functional significance of these RNAs [1, 2]. However, the profile of chromosome-associated small RNAs in mammals has not yet been systematically investigated. In this study, we performed deep sequencing of chromosome-associated small RNAs in the metaphase of mouse 3T3 cell lines under three different conditions (buffer A, low salt, and high salt), as previous research indicates that different concentrations of salt solutions can extract various types of small RNAs associated with the metaphase chromosomes while maintaining the stability of the core chromosome structure. We aim to identify small RNAs tightly bound to the metaphase chromosomes across the three conditions, which indicates the functional significance of these RNAs during mitosis and possibly in disease mechanisms [3].

Methods and materials: Mouse 3T3 cells were cultured in DMEM supplemented with 10 % heat inactivated fetal calf serum and penicillin-streptomycin. $5X10^7$ Cells were harvested when they achieved 70-80 % confluence and were treated with colcemid (100 ng/ml) for 12 hours to synchronized at the metaphase.

The mitotic chromosomes were isolated as previously described in Meng et al 2016 [1]. In addition to the low salt (LS) and high salt (HS) samples, a third set of chromosomes (buffer A (BA)sample) was treated with Buffer A (15 mM Tris–HCl (pH 7.5), 80 mM KCl, 2 mM EDTA, 2 mM spermine, 5 mM spermidine, 0.1 mM PMSF, and 0.05 % digitonin) on ice for 3 minutes. The mixture was centrifuged at 1750g for 6 minutes and then resuspended in the PA buffer before loading onto the sucrose gradient. [interphase nuclear miRNA extraction] Total RNA was extracted from each sample with TRIzol according to the manufacturer's instruction (Thermo Fisher Scientific). The small RNA library was prepared with TruSeq® Small RNA Library Preparation Kits following the Illumina's protocol. Deep sequencing of the library was performed on Illumina HiSeq platform using single end 1x50 as the sequencing mode.

Raw sequencing data were first processed with FASTX-Toolkit to remove sequencing adaptors and low-quality reads (http://hannonlab.cshl.edu/fastx_toolkit/). Reads longer than 30 bp accounted for a very small part of all reads due to the size selection of gel purification in the library preparation and were removed from downstream analysis. Bowtie was then used to map the clean reads generated from above steps to the mouse genome (version mm10) with default parameters except -v -k and -m all set to 1. Only uniquely mapped reads were used for downstream analysis.

Reads were grouped into the same cluster if the distance between their 5' ends were less than 15 bp, which is the minimal length of annotated mouse miRNAs. In this study, the highest-frequency read in each read cluster was used to defined the position of the read cluster.

For each condition (BA, LS or HS), read clusters were ranked by their RPM and a cut-off of RPM was set such that the total number of reads in clusters with RPM above the cut-off accounted for >90 % of all reads in the condition. The read clusters above the cut-off were then compared with the mouse genomic features (version mm10). The microRNA (miRNA) annotation data were retrieved from miRBase (miRBase v21; http://www.mirbase.org/) [4]. The Piwi-interacting RNA (piRNA) and transcription start site (TSS) annotation data were retrieved from UCSC genome browser mouse mm10. Read clusters were identified as miRNAs or piRNAs if they fell into the annotated genomic regions of miRNA or piRNAs, respectively.

miREval2.0 (http://mimirna.centenary.org.au/mireval/) were employed to predict novel miRNAs from read clusters above the RPM cut-off that does not match any annotated small RNA [5]. The express sequence tag (EST) annotation used to support the predicted novel miRNAs were also retrieved from UCSC genome browser mouse mm10.

Results: Here we only used the uniquely mapped reads for downstream data analysis. We grouped reads into the same cluster if the distance between their 5' ends were less than 15 bp (Methods and Materials). By doing this, we obtained 122940, 77275, 75730, 96720 read clusters for BA, LS, HS and NC, respectively. Here we selected read clusters with large read numbers and made comparisons between the three conditions. For each condition, we first set a cut-off for the read numbers such that the total number of reads in clusters above the cut-off accounted for at least 90% of all reads in this condition. Furthermore, to make quantitative comparisons between conditions, we used reads per million (RPM) to normalize the read numbers of each read cluster (Methods and Materials). In the following paragraphs, we used RPM to measure the amount of small RNAs associated with chromosomes.

For those read clusters above the cut-offs, we compared the RPM distributions between the three conditions. We found that, in BA and LS, the average RPM was very close to each other, while the average RPM in HS was significantly lower than in the other two conditions. This result shows that small RNAs of high concentration are more likely to be washed away from chromosomes in high-salt environment.

The read clusters above the cut-offs are highly consistent among the three conditions. We picked the top-ranked (by RPM) 148 read clusters in each condition and found 128 read clusters in common, demonstrating a statistically significant overlap among them [6].

The three sets have a common set of 132 read clusters whose association with metaphase chromosomes are both high and stable (i.e., not significantly influenced by salt concentration), suggesting their potential functional importance in mitosis. By contrast, most of the 132 clusters have a relatively large variation in RPM between condition

HS/LS/BA and condition NC (i.e., the interphase cell nucleus of the mouse 3T3 cell line). Specifically, some of the 132 clusters have very low RPMs in condition NC, indicating their special roles in metaphase.

Based on the experimental protocols and small RNA annotations, we classified the majority of 132 read clusters mentioned above into three categories: microRNA (miRNA), Piwi-interacting RNA and TSS-associated (TSSa) RNA. We first compared the 132 read clusters with annotated mouse miRNAs and found that 72 of 132 could perfectly match the miRNA genomic locations. Thus, the 72 read clusters were classified into annotated mouse miRNAs. As we expected, in all three conditions, the total reads of the 72 miRNAs accounted for more than half of the total reads of the 132 clusters, but the average read numbers showed no significant difference between the 72 miRNAs and the rest.

Piwi-interacting RNA (piRNA) constitutes the largest category of non-coding small RNAs in mammals. We compared the remaining 60 read clusters (out of 132) with annotated mouse piRNAs and found 6 read clusters to be piRNAs. For the rest (54) of the 132 read clusters, there were 6 of them located in the vicinity of transcription start sites (TSSs), making it highly possible that the 6 read clusters belonged to the category of TSSa RNA. The rest (48) of the 132 read clusters could not be mapped to any known small RNA; since they all had very high abundance (comparable with the 72 miRNAs), they were likely to be unknown chromosome-associated small RNAs.

Conclusion: In this study, we performed deep sequencing of chromosome-associated small RNAs in the metaphase of mouse 3T3 cell lines under three different conditions: buffer A, low salt, and high salt. The small RNAs identified with high abundance were consistent across the three conditions, indicating the functional significance of these stably attached RNAs. Most of these small RNAs can be mapped to known features such as miRNAs and piRNAs, while a considerable amount (36.3 %) could not be annotated. Additionally, some unannotated small RNAs are not present in interphase 3T3 cell lines, suggesting their possible mitosis-specific functions. Analyzing these different types of small RNAs helps us better understand mitosis, leading to new insights into the mechanisms of cell division and possible role in pathogenesis.

Funding: This work was supported by grants from the National Key R&D Program of China (2020YFA0908100), the National Natural Science Foundation of China (32370572) and the K.C. Wong Education Foundation (H.K.).

References

1. Meng Y., Yi X., Li X., Hu C., Wang J., Bai L., Czajkowsky D. M., Shao Z. The non-coding RNA composition of the mitotic chromosome by 5'-tag sequencing. *Nucleic Acids Res*. 2016;44(10):4934-4946. doi 10.1093/nar/gkw195
2. Zhang L., Hu C., Xu Z., Li H., Ye B., Li X., Czajkowsky D.M., Shao Z. Quantitative catalogue of mammalian mitotic chromosome-associated RNAs. *Sci Data*. 2024;11(1):43. doi 10.1038/s41597-023-02884-8
3. Orlov Y.L., Chen W.L., Sekacheva M.I., Cai G., Li H. Editorial: high-throughput sequencing-based investigation of chronic disease Markers and mechanisms. *Front Genet.* 2022;13:922206. doi 10.3389/fgene.2022.922206
4. Kozomara A., Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 2014;42(D1):D68-D73. doi 10.1093/nar/gkt1181
5. Gao D., Middleton R., Rasko J.E., Ritchie W. miREval 2.0: a web tool for simple microRNA prediction in genome sequences. *Bioinformatics.* 2013;29(24):3225-3226. doi 10.1093/bioinformatics/btt545
6. Li H., Tong P., Gallegos J., Dimmer E., Cai G., Molldrem J.J., Liang S. PAND: A distribution to identify functional linkage from networks with preferential attachment property. *PLoS One.* 2015;10(7): e0127968. doi 10.1371/journal.pone.0127968

1 Симпозиум «Геномика, транскриптомика и биоинформатика»
Symposium “Genomics, transcriptomics and bioinformatics”

doi 10.18699/bgrs2024-1.2-01

Роль некодирующих РНК в эпигенетической трансформации генных локусов

Беспалова А.В.*, Куликова Д.А., Фуников С.Ю.
Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия
** lina.bespalowa@yandex.ru*

Ключевые слова: парамутация; piРНК; эпигенетика; дрозофила; генные piРНК

Мотивация и цель: Феномен парамутации описывает взаимодействие между двумя аллелями, при котором один аллель инициирует наследуемые эпигенетические изменения в другом аллеле, не затрагивая последовательность ДНК [1]. Эпигенетическое преобразование вследствие парамутации заключается в изменении паттерна метилирования гистонов и/или ДНК парамутабельного аллеля и оказывает влияние на экспрессию генов. Механизм эпигенетического преобразования аллелей при парамутации остается малоизученным, однако все ранее описанные случаи объединяет то, что главными эффекторными молекулами, необходимыми для инициации и поддержания эпигенетических модификаций, являются короткие некодирующие РНК [1].

В основе парамутации у дрозофил лежит биохимический путь образования piРНК (короткие некодирующие РНК, ассоциированные с белком Piwi) и механизм котранскрипционного сайленсинга с помощью комплекса piРНК/Piwi [2]. Образование piРНК происходит исключительно в репродуктивных органах животных и направлено на подавление активности мобильных элементов. Однако в редких случаях в ходе эпигенетического сайленсинга мобильных элементов белковый комплекс биогенеза piРНК может ошибочно нацеливаться на белок-кодирующие гены и подавлять их экспрессию [3, 4]. У *D. virilis* piРНК-опосредованное подавление аллеля *cdi* является примером подобной “непреднамеренной” регуляции [5]. Кроме того, материнское наследование piРНК к транскриптам *cdi* приводит к эпигенетической трансформации гомологичного локуса вследствие парамутации [6].

В нашем исследовании мы изучили репертуар генных piРНК у линий *D. virilis* из различных географических и лабораторных популяций, исследовали их многообразия и пути возникновения, а также изучили, может ли материнское наследование piРНК комплементарных мРНК белок-кодирующих генов приводить к эпигенетической трансформации генных локусов в результате парамутации.

Методы и алгоритмы: Для анализа генных piРНК мы использовали 11 линий *D. virilis*, представляющих природные и лабораторные популяции. Библиотеки коротких некодирующих РНК приготовлены с помощью набора Truseq small RNA kit (Illumina, США) из тотальной РНК, выделенной из 30 пар яичников самок каждой из линий. Удаление фракции 2S рибосомальной РНК перед созданием библиотеки проведено с помощью выделения 18–30 нт фракции тотальной РНК в полиакриламидном геле. Библиотеки секвенированы на приборах Nextseq-2000. Анализ экспрессии piРНК выполнен с использованием собственного биоинформационно-вычислительного конвейера, написанного на языке R.

Исследование парамутагенных свойств piРНК-продуцирующих локусов выполнено на модели скрещиваний между самками линии *140* и самцами линии *9*. Сначала были отобраны самки линии *140*, содержащей потенциальные парамутагенные аллели *Adar/OtopLb*$^{140/140}$ и *RhoGEF3*$^{140/140}$, и самцы линии *9*, аллели которой являются парамутабельными по отношению к аллелям линии *140*. Затем были поставлены скрещивания самок линии *140* с самцами линии *9* (♀*140* x ♂*9*) и получены гибриды первого поколения (F1). Далее из гибридов F1 были отобраны девственные самки и проведены возвратные скрещивания с самцами исходной родительской линии *9* для получения гомозиготного потомства по локусам *Adar*/*OtopLb* и *RhoGEF3* линии *9*. Генотипирование выведенных линий, гомозиготных по аллелям *Adar*$^{9/9}$ и *RhoGEF3*$^{9/9}$, выполнено с помощью аллель-специфичных праймеров. Только потомство от самок с генотипом *Adar*$^{9/9}$ и *RhoGEF3*$^{9/9}$ было использовано для получения библиотек коротких некодирующих РНК.

Результаты: Среди 11 линий *D. virilis* из разных географических и лабораторных популяций мы обнаружили более 100 белок-кодирующих генов, на мРНК которых картируется piРНК с уровнем экспрессии не менее 50 rpm (reads per million, количество прочтений на миллион выровненных прочтений) хотя бы в одной из исследованных линий. Мы обнаружили, что линии *D. virilis* обладают выраженным полиморфизмом экспрессии генных piРНК. Из совокупности обнаруженных генов только для 25 последовательностей образование piРНК наблюдается более чем в 5 линиях, а высокоэкспрессирующиеся генные piРНК, с покрытием более 500 rpm, обнаружены только в одной из линий – 160.

Далее мы отобрали гены с наибольшим количеством картированных piРНК; были выбраны три гена – *Adar*, *OtopLb* и *RhoGEF3*, на последовательности которых в *140* линии уникально картируется более 500 rpm piРНК, а в *9* линии piРНК отсутствуют. Мы обнаружили, что материнское наследования piРНК к *Adar*, *OtopLb* и *RhoGEF3* определяет экспрессию piРНК из данных локусов в следующем поколении и наблюдается только у гибридов ♀*140* × ♂*9*, когда piРНК наследуются по материнской линии, как и в случае гена *cdi* [6].

Взаимосвязь между образованием генных piРНК и экспрессией мРНК генов не всегда очевидна. Между уровнем экспрессии piРНК к последовательности гена *Adar* и уровнем его экспрессии в *140* и *9* линии, а также гибридов от прямого и реципрокного скрещивания наблюдается отрицательная корреляция. Однако для гена *OtopLb*, который находится в одном локусе с геном *Adar*, мы, наоборот, обнаружили положительную корреляцию между экспрессией мРНК и piРНК в линии *140*. Для гена *RhoGEF3* не удалось установить достоверную корреляцию между экспрессией piРНК и уровнем экспрессии гена. Гены *Adar* и *OtopLb* находятся в одном локусе в области прицентромерного гетерохроматина, а ген *RhoGEF3* на другой хромосоме также в гетерохроматиновой области. Проведя сравнительный анализ строения генных локусов, мы обнаружили, что в локус гена *RhoGEF3* в линии *140*, в отличие от *9*, произошла вставка ретротранспозона *Nausicaa*. Локусы генов *Adar* и *OtopLb* в линиях *140* и *9*, за исключением мелких перестроек, гомологичны последовательности в референсном геноме.

Чтобы проверить, может ли материнское наследование piРНК, комплементарных транскриптам *Adar*, *OtopLb* и *RhoGEF3,* вызывать трансформацию гомологичных геномных локусов в piРНК кластеры в результате парамутации, мы провели скрещивания между самками линии *140* и самцами линии *9*. В результате эксперимента мы вывели 16 линий с генотипом *Adar/OtopLb*$^{9/9}$ и *RhoGEF3*$^{9/9}$.

Однако проведенный анализ данных линий показал отсутствие экспрессии piРНК из локусов *Adar/OtopLb* и *RhoGEF3* во всех изученных нами линиях.
Выводы: В данной работе мы провели исследование генных piРНК в 11 линиях *D. virilis* из различных географических и лабораторных популяций. Генные piРНК составляют примерно 5 % от общего пула piРНК в яичниках и обладают выраженным межлинейным полиморфизмом экспрессии. Мы установили, что многообразие генных piРНК у *D. virilis* определяется различными механизмами образования генных piРНК. Используя модель скрещиваний линии *D. virilis 9* и *140*, демонстрирующих асимметричную экспрессию piРНК из локусов генов *Adar*, *Otoplb* и *RhoGEF3*, мы установили, что экспрессии и материнского наследования piРНК к транскриптам генов *Adar*, *OtopLb* и *RhoGEF3* недостаточно для формирования piРНК-продуцирующих кластеров в гомологичных локусах. Мы также обнаружили инсерцию мобильного элемента *Nausicaa* в локусе *RhoGEF3* в линии *140*. Учитывая, что инсерции транспозонов отсутствуют в этой же области локуса *RhoGEF3* в линии *9*, мы предполагаем, что именно инсерция мобильного элемента является детерминантой образования piРНК кластера в локусе *RhoGEF3* в линии *140*.
Финансирование: Работа выполнена при финансовой поддержке гранта Российского научного фонда (проект № 22-74-10050).

The role of non-coding RNAs in the epigenetic conversion of gene loci

Bespalova A.V.*, Kulikova D.A., Funikov S.Y.
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
** lina.bespalowa@yandex.ru*

Key words: paramutation; Piwi-interacting RNAs; epigenetics; drosophila; genic piRNAs

Motivation and Aim: The phenomenon of paramutation describes the interaction between two alleles, in which one allele initiates inherited epigenetic changes in the other allele without affecting the DNA sequence [1]. Epigenetic conversion due to paramutation involves changes in methylation pattern of histones and/or DNA of a paramutable allele, and affects gene expression. The mechanism of epigenetic conversion of alleles during paramutation remains poorly understood, however, all previously described cases are united by the fact that the main effector molecules necessary for the initiation and maintenance of epigenetic modifications are small non-coding RNAs 1].
Paramutation in drosophila is based on the biochemical pathway of piRNAs (short non-coding RNAs associated with the Piwi protein) and the mechanism of co-transcriptional silencing by piRNA/Piwi complex [2]. The biogenesis of piRNAs occurs exclusively in the reproductive organs of animals and is aimed at suppressing the activity of transposable elements. However, in rare cases, during epigenetic silencing of transposons, the protein complex of piRNA biogenesis may mistakenly target protein-coding genes and suppress their expression [3, 4]. In *D. virilis*, rRNA-mediated suppression of the *cdi* allele is an example of such “unintended” regulation [5]. In addition, maternal inheritance of piRNA to *cdi* transcripts leads to epigenetic transformation of the homologous locus due to paramutation [6].

Herein, we analyzed the repertoire of genic piRNAs in *D. virilis* strains from various geographical and laboratory populations, investigated their diversity and pathways of origin, and also studied whether maternal inheritance of piRNAs targeted mRNA of protein-coding genes can lead to epigenetic conversion of gene loci as a result of paramutation.

Methods and Algorithms: To analyze genic piRNAs, we used 11 *D. virilis* strains representing natural and laboratory populations. Libraries of small non-coding RNAs were prepared using the Truseq small RNA kit (Illumina, USA) from total RNA isolated from 30 pairs of ovaries of females of each strain. The removal of the 2S ribosomal RNA fraction before the cloning of the library was carried out by isolating the 18–30 nt fraction of total RNA in a polyacrylamide gel. The libraries were sequenced on Nextseq-2000 platform. The analysis of piRNA expression was performed using our own pipeline written in the R language.

The study of the paramutagenic properties of piRNA-producing loci was performed on a model of crosses invloving females of strain *140* and males of strain *9*. First, females of strain *140* containing potential paramutagenic alleles $Adar/OtopLb^{140/140}$ and $RhoGEF3^{140/140}$ and males of strain *9*, whose alleles are paramutable with respect to alleles of strain *140*, were selected. Then crosses of females of strain *140* with males of strain *9* (♀ *140* × ♂ *9*) were performed and hybrids of the first generation (F1) were obtained. Further, virgin females were selected from F1 hybrids and recurrent crosses with males of the original parental strain *9* were performed to obtain homozygous offspring at the *Adar/OtopLb* and *RhoGEF3* strain *9* loci. Genotyping of the derived homozygous strains for the $Adar^{9/9}$ and $RhoGEF3^{9/9}$ alleles was performed using allele-specific primers. Only offspring from females with the $Adar^{9/9}$ and $RhoGEF3^{9/9}$ genotypes were used to obtain libraries of small non-coding RNAs.

Results: Among 11 *D. virilis* strains from different geographical and laboratory populations, we found more than 100 protein-coding genes, on the mRNA of which piRNA is mapped with an expression level of at least 50 rpm (reads per million, the number of reads per million aligned reads) in at least one of the studied lines. We found that *D. virilis* strain have a pronounced polymorphism of genic piRNA expression. Of the total of the detected genes, only for 25 sequences, the formation of piRNAs is observed in more than 5 lines, and highly expressed genic piRNAs (more than 500 rpm) were found in only one of the strains – *160*.

Next, we selected the genes with the largest number of mapped piRNAs. Three genes were selected – *Adar*, *OtopLb* and *RhoGEF3*, on the sequence of which more than 500 rpm piRNAs are uniquely mapped in strain *140*, and there are no piRNAs in strain *9*. We found that maternal inheritance of piRNA to *Adar*, *OtopLb* and *RhoGEF3* determines the expression of piRNA from these loci in the next generation and is observed only in hybrids *140×9* when piRNAs are inherited maternally, as in the case of the *cdi* gene [6].

The relationship between the formation of genic piRNAs and the expression of mRNA is not always obvious. There is a negative correlation between the level of expression of piRNA to the sequence of the *Adar* gene and the level of its mRNA in strains *140* and *9*, as well as hybrids from direct and reciprocal crossing. However, for the *OtopLb* gene, which is located at the same locus with the *Adar* gene, on the contrary, we found a positive correlation between the expression of mRNA and piRNA in strain *140*. For the *RhoGEF3* gene, it was not possible to establish a reliable correlation between the expression of piRNA and the level of gene expression. The *Adar* and *OtopLb* genes are

located at the same locus in the region of the centromeric heterochromatin, and the *RhoGEF3* gene on the other chromosome is also in the heterochromatin region. After conducting a comparative analysis of the structure of the gene loci, we found that the retrotransposon *Nausicaa* was inserted into the locus of the *RhoGEF3* gene in strain *140*, in contrast to strain *9*. The loci of the *Adar* and *OtopLb* genes in *140* and *9* strains, with the exception of minor rearrangements, are homologous to the sequence in the reference genome.

To test whether maternal inheritance of piRNAs complementary to the *Adar*, *OtopLb* and *RhoGEF3* transcripts can cause the conversion of homologous genomic loci into piRNA clusters as a result of paramutation, we performed crosses between females of strain *140* and males of strain *9*. As a result of the experiment, we obtained 16 strain with the genotype *Adar/OtopLb*$^{9/9}$ and *RhoGEF3*$^{9/9}$. However, the analysis of these strains showed the absence of expression of piRNA from the *Adar/OtopLb* and *RhoGEF3* loci in all the obtained so far.

Conclusion: In this work, we performed a study of genic piRNAs in 11 *D. virilis* strains from various geographical and laboratory populations. Genic piRNAs make up approximately 5 % of the total pool of piRNAs in the ovaries and have a pronounced intrastrain polymorphism of expression. We have established that the diversity of genic piRNAs in *D. virilis* is determined by various mechanisms of their formation. Using the crosses involving strains *9* and *140*, demonstrating asymmetric expression of piRNA from the loci of the *Adar*, *Otoplb* and *RhoGEF3* genes, we found that the expression and maternal inheritance of piRNA is not sufficient for the formation of piRNA clusters in homologous loci. We also found the insertion of the transposable element *Nausicaa* at the *RhoGEF3* locus in strain *140*. Considering that there are no transposon insertions in the same region of the *RhoGEF3* locus in strain *9*, we assume that transposon insertion is determining the formation of a piRNA cluster at the *RhoGEF3* locus in strain *140*.

Funding: The study is supported by Russian Science Foundation (grant No. 22-74-10050).

Список литературы/References

1. Hollick J.B. Paramutation and related phenomena in diverse species. *Nat Rev Genet*. 2017(1):5-23. doi 10.1038/nrg.2016.115
2. Czech B., Munafò M., Ciabrelli F., Eastwood E.L., Fabry M.H., Kneuss E., Hannon G.J. piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet.* 2018;52:131-157. doi 10.1146/annurev-genet-120417-031441
3. Blumenstiel J.P., Erwin A.A., Hemmer L.W. What Drives Positive Selection in the *Drosophila* piRNA Machinery? The Genomic Autoimmunity Hypothesis. *Yale J Biol Med.* 2016;89(4):499-512
4. Rozhkov N.V., Aravin A.A., Zelentsova E.S., Schostak N.G., Sachidanandam R., McCombie W.R., Hannon G.J., Evgen'ev M.B. Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *RNA*. 2010(8):1634-1645. doi 10.1261/rna.2217810
5. Erwin A.A., Galdos M.A., Wickersheim M.L., Harrison C.C., Marr K.D., Colicchio J.M., Blumenstiel J.P. piRNAs are associated with diverse transgenerational effects on gene and transposon expression in a hybrid dysgenic syndrome of *D. virilis*. *PLoS Genet*. 2015;11(8):e1005332. doi 10.1371/journal.pgen.1005332
6. Dorador A.P., Dalikova M., Cerbin S., Stillman C.M., Zych M.G., Hawley R.S., Miller D.E., Ray D.A., Funikov S.Y., Evgen'ev M.B., Blumenstiel J.P. Paramutation-like epigenetic conversion by pirna at the telomere of *Drosophila virilis*. *Biology (Basel)*. 2022;11(10):1480. doi 10.3390/biology11101480

doi 10.18699/bgrs2024-1.2-02

Транскриптомика единичных клеток позволяет считать *Acomys cahirinus* новой перспективной моделью для изучения надпочечников

Быданов А.[1]*, Билялова А.[1], Билялов А.[1, 2], Газизова Г.[1], Шагимарданова Е.[1], Козлова О.[1], Гусев О.[1, 3]

[1] *Исследовательский центр регуляторной геномики, Институт фундаментальной медицины и биологии, Казанский федеральный университет, Казань, Россия*

[2] *ГБУЗ Московский клинический научный центр им. А.С. Логинова ДЗМ, Москва, Россия*

[3] *Высшая медицинская школа, Университет Джунтендо, Токио, Япония*

* *bydanovandrey03@gmail.com*

Ключевые слова: *Acomys cahirinus*; транскриптомика единичных клеток; надпочечник; *Cyp17a1*

Мотивация и цель: Изучение заболеваний надпочечника в наше время является актуальной проблемой. В свете разработки генетических и клеточных методов лечения заболеваний надпочечника крайне важно проводить доклинические исследования с использованием животных моделей, таких как мыши и крысы. Тем не менее между лабораторными животными и человеком существуют заметные различия в морфологических и генетических аспектах надпочечников. В связи с этим возникает вопрос о целесообразности использования этих животных в качестве моделей для изучения заболеваний надпочечника, а также для оценки новых методов лечения. *Acomys cahirinus* – вид, широко известный как иглистая мышь, является перспективным модельным организмом для изучения различных физиологических процессов благодаря своим замечательным способностям к регенерации и сходству с человеком в некоторых биологических аспектах, особенно в отношении надпочечников. В наших предыдущих исследованиях мы уже продемонстрировали морфологически сходную картину надпочечника человека и иглистой мыши. Цель данного исследования – изучить структуру надпочечника *Acomys cahirinus* с разрешением на уровне отдельных клеток.
Методы и алгоритмы: Мы использовали протокол 10X Genomics Single Cell 5' R2-only [1] для секвенирования РНК надпочечников 10 образцов *Acomys cahirinus* – 5 самцов и 5 самок. Все образцы были закартированы на геном *Acomys cahirinus* в программе cellranger count, что дало нам около 88 % картирования и 73 % уникально картированных прочтений. Дальнейшая обработка проводилась с помощью стандартного пайплайна Seurat V5 [2] с нормализацией SCT и интеграцией с помощью Harmony [3].
Результаты: Контроль качества чтений 10 образцов надпочечника проводился по стандартным критериям. Интегрированный набор данных был проанализирован с использованием маркеров для надпочечников мыши, что позволило выявить следующие типы клеток: клубочковая, сетчатая и пучковая зоны, мозговое вещество, иммунные клетки, эндотелиальные клетки, нейроглия и мезенхимальные клетки. Интеграция с образцами лабораторной мыши *Mus musculus* показала значительную разницу в составе типов клеток, а именно: среди них были выявлены почти все типы клеток *Acomys*, за исключением сетчатой

зоны. Кроме того, в клетках *Acomys* был выявлен значительный уровень экспрессии гена *Cyp17a1*, который не экспрессировался в клетках мыши, но известен своей экспрессией в надпочечниках человека.
Выводы: В заключение следует отметить, что наше исследование позволяет рассматривать *Acomys cahirinus* как новую и надежную модель для изучения заболеваний надпочечника. Сочетание морфологического и транскриптомного подходов обеспечивает комплексную основу для понимания биологии надпочечника и разработки целевых методов лечения их заболеваний. Данное исследование способствует расширению наших знаний о функции и патофизиологии надпочечника, что в конечном итоге позволит улучшить диагностику и лечение заболеваний надпочечника как у человека, так и у других видов млекопитающих.
Финансирование: Исследование поддержано грантом Министерства науки и высшего образования Российской Федерации 075-15-2021-1344.

Single-cell transcriptomics suggests *Acomys cahirinus* to be a new promising model for adrenal gland studies

Bydanov A.[1]*, Bilyalova A.[1], Bilyalov A.[1, 2], Gazizova G.[1], Shagimardanova E.[1], Kozlova O.[1], Gusev O.[1, 3]

[1] *Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[2] *Moscow Clinical Research Center named after A.S. Loginov MHD, Moscow, Russia*

[3] *Graduate School of Medicine, Juntendo University, Tokyo, Japan*

* *bydanovandrey03@gmail.com*

Key words: *Acomys cahirinus*; scRNAseq; adrenal gland; *Cyp17a1*

Motivation and Aim: The study of adrenal gland disorders is a pressing issue in the present era. In light of the development of genetic and cellular therapies for adrenal conditions, it is imperative that preclinical research using animal models, such as mice and rats, be conducted. Nevertheless, there are notable differences in the morphological and genetic aspects of the adrenal gland between laboratory animals and humans. This prompts the question of whether it is appropriate to use these animals as models for studying adrenal disorders, and indeed for evaluating new treatments. *Acomys cahirinus*, commonly known as the spiny mouse, has emerged as a promising model organism for investigating various physiological processes, due to its remarkable regenerative capabilities and similarities to humans in certain biological aspects, particularly with regard to the adrenal gland. In our previous research, we have already demonstrated a morphologically analogous picture of the human and spiny mouse adrenal glands. The objective of this study is to investigate the structure of the *Acomys cahirinus* adrenal gland at the single-cell resolution.
Methods and Algorithms: We utilized 10X Genomics Single Cell 5' R2-only protocol [1] for RNA sequencing of adrenal glands of 10 *Acomys cahirinus* samples – 5 male and 5 female. All samples were mapped onto *Acomys cahirinus* genome with cellranger count tool, that gave us approximately 88 % mapping rate and 73 % of reads uniquely

mapped. Further processing was carried out with standard Seurat V5 [2] pipeline with SCT normalization and integration using Harmony [3].

Results: A quality control of reads from 10 samples was performed using standard criteria. The integrated dataset was analyzed using markers for mouse adrenal gland, which helped us to reveal the following cell types: zona reticularis, zona fasciculata, zona glomerulosa, medulla, immune cells, endothelial cells, neuroglia, and mesenchymal cells. The integration with *Mus musculus* showed a significant difference in cell types composition, with the mouse dataset exhibiting almost all *Acomys* cell types, except for zona reticularis. Furthermore, we observed a significant expression level of *Cyp17a1* gene in *Acomys* cells, which exhibited no expression in mouse cells but is known for its expression in human adrenal gland.

Conclusion: In conclusion, our study establishes *Acomys cahirinus* as a novel and robust model for investigating adrenal gland diseases. The combination of morphological and transcriptomic approaches provides a comprehensive framework for understanding adrenal gland biology and developing targeted therapies for adrenal disorders. This research contributes to advancing our knowledge of adrenal gland function and pathophysiology, ultimately paving the way for improved diagnosis and treatment of adrenal gland diseases in both humans and other mammalian species.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation grant 075-15-2021-1344.

Список литературы/References

1. https://cdn.10xgenomics.com/image/upload/v1666737555/support-documents/CG000331_ChromiumNextGEMSingleCell5-v2_UserGuide_RevE.pdf
2. Hao Y., Stuart T., Kowalsk M.H. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. 2024;42:293-304. doi 10.1038/s41587-023-01767-y
3. Korsunsky I., Millard N., Fan J., Slowikowski K., Zhang F., Wei K., Baglaenko Y., Brenner M., Loh P.R., Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;(12):1289-1296. doi 10.1038/s41592-019-0619-0

doi 10.18699/bgrs2024-1.2-03

Цис-элементы, регулирующие экспрессию генов в ответ на ключевые фитогормоны у *Arabidopsis thaliana* L.

Долгих В.А.[1, 2]*, Землянская Е.В.[1, 2], Левицкий В.Г.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *dolgikh@bionet.nsc.ru*

Ключевые слова: регуляция транскрипции; *цис*-элементы; транскрипционные факторы

Мотивация и цель: Транскрипционные факторы (ТФ) управляют экспрессией генов-мишеней, связываясь с короткими участками в промоторных областях, называемых цис-элементами. Последовательности *цис*-элементов, с которыми связывается определенный ТФ, как правило, являются частично вырожденными, и их спектр обобщается в виде мотивов. Достаточно часто для эффективной работы ТФ необходимо формирование гетеродимеров с ТФ-партнерами. В зависимости от ТФ в составе комплекса гетеродимер может садиться на ДНК в результате взаимодействия только одного из ТФ с соответствующим *цис*-элементом или обоих ТФ с так называемым композиционным элементом, состоящим из двух *цис*-элементов, которые могут быть разделены спейсером или частично перекрываться друг с другом. Из-за комбинаторной сложности транскрипционной регуляции ее экспериментальные исследования часто ограничиваются установлением отдельных *цис*-элементов в регуляторных районах отдельных генов. Альтернативным подходом, дающим возможность системного выявления и функционального исследования *цис*-элементов, является совместный анализ полногеномных данных разного типа, описывающих разные аспекты транскрипционной регуляции экспрессии генов (например, ChIP-seq и RNA-seq). Потенциал такого подхода не реализован полностью, и потому создание соответствующего биоинформатического инструментария и его использование для системного изучения регуляции экспрессии генов являются актуальной задачей, к решению которой мы обратились в данной работе.

Методы и алгоритмы: Мы разработали конвейер SFmotif (сокращение от structure and function of motif), предназначенный для совместного анализа данных ChIP-seq и RNA-seq с целью комплексного изучения механизмов регуляции транскрипции. Он состоит из трех модулей, позволяет извлекать обогащенные в пиках ChIP-seq мотивы и оценивать их влияние на транскрипционную активность генов. Конвейер реализован в виде R-пакета и запускается из командной строки Linux. SFmotif интегрирует как существующие программные решения (например, Homer [1], MCOT [2]), так и оригинальные функции и модули, написанные на языке R. ChIP-seq данные по связыванию ТФ ABF1, BES1, MYC2 и ARR10, а также индуцированные этими гормонами транскриптомы были взяты из открытых источников [3, 4].

Результаты: Мы применили разработанный нами конвейер SFmotif для исследования *цис*-элементов, регулирующих транскрипционный ответ генов на четыре ключевых фитогормона: абсцизовую кислоту, брассиностероиды, жасмонаты и цитокинин. Активация их сигнальных путей приводит к активации

небольшого числа специфических ТФ, однако позволяет запускать разнообразные транскрипционные ответы в зависимости от условий. Это подразумевает наличие дополнительных механизмов, модулирующих активность соответствующих ТФ, и делает актуальной задачу их исследования. Используя ChIP-seq данные по связыванию ABF1, BES1, MYC2 и ARR10 (которые регулируют ответ на абсцизовую кислоту, брассиностероиды, жасмонаты и цитокинин соответственно), а также индуцированные этими гормонами транскриптомы, нам удалось выявить обогащение мотивов, схожих с известными сайтами связывания этих ТФ. Все они были ассоциированы с повышением экспрессии генов в ответ на действие гормонов. Помимо этих мотивов, в пиках ChIP-seq были обогащены мотивы, соответствующие, вероятно, сайтам связывания потенциальных ТФ-партнеров. Для ТФ ABF1 и MYC2 были обнаружены функциональные композиционные элементы, с которыми, вероятно, эти ТФ могут связываться в составе гетеродимеров с ТФ-партнерами. Примечательно, что классический сайт связывания ТФ ARR10 не является наиболее обогащенным, в то же время для ARR10 обнаружено наибольшее количество мотивов для потенциальных ТФ-партнеров по сравнению с другими ТФ. Это может свидетельствовать о том, что функциональность ТФ ABF1, BES1 и MYC2 регулируется в первую очередь нуклеотидным контекстом сайта связывания, а в активности ARR10 принципиальную роль играют ТФ-партнеры.

Выводы: Разработан программный конвейер SFmotif для комплексного изучения механизмов регуляции активности генов. Предсказано, что фитогормоны могут реализовывать разные стратегии транскрипционной регуляции с различным вкладом транскрипционных факторов партнеров.

Финансирование: Исследование поддержано грантом РНФ № 20-14-00140.

Cis-elements regulating gene expression in response to the key phytohormones in *Arabidopsis thaliana* L.

Dolgikh V.[1, 2]*, Zemlyanskaya E.[1, 2], Levitsky V.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *dolgikh@bionet.nsc.ru*

Key words: transcription regulation; *cis*-elements; transcription factors

Motivation and Aim: Transcription factors (TFs) control the expression of their target genes by binding to the short sequences in promoter regions called *cis*-elements. *cis*-elements bound by a certain TF usually have degenerate nucleotide seqiences, and their whole set is usually generalized as a motif. Quite often, an effective TF functioning requires formation of heterodimers with partner TFs. Depending on the TFs in the complex, the heterodimer can bind to DNA due to an interaction of only one TF with the corresponding *cis*-element or both TFs with the so-called composite element consisting of two *cis*-elements, which can be either separated by a spacer or partially overlap with each other. Due to the combinatorial complexity of the transcriptional regulation, its experimental studies are often limited to the discovery of individual *cis*-elements in the

regulatory regions of individual genes. Alternatively, an integrated analysis of different genome-wide data, which describe different aspects of transcriptional regulation of gene expression (for example, ChIP-seq and RNA-seq) makes systematic identification and functional studies of *cis*-elements possible. The potential of this approach is far from being thoroughly realized, and therefore the development of appropriate bioinformatics tools and and their use for the systematic studies of the regulation of gene expression is a relevant task, which we have addressed in this work.

Methods and Algorithms: We have developed the SFmotif pipeline (short for "structure and function of motif"), which was designed for the integrated analysis if ChIP-seq and RNA-seq data to comprehensively study the mechanisms of transcriptional regulation. It consists of three modules and allows extracting the motifs enriched in ChIP-seq peaks and evaluate their influence on the transcriptional activity of genes. The pipeline is implemented as an R package and it is launched from the Linux command line. SFmotif integrates both existing software solutions (e. g. Homer [1], MCOT [2]), as well as original functions and modules written in the R language. ChIP-seq data on the binding of TFs ABF1, BES1, MYC2 and ARR10, as well as the transcriptomes induced by these hormones were taken from publicly available repositories [3, 4].

Results: We used SFmotif pipeline to study *cis*-elements, which that regulate the transcriptional response of genes to four key phytohormones: abscisic acid, brassinosteroids, jasmonates, and cytokinin. Activation of their signaling pathways leads to the activation of a limited number of specific TFs, which, however, trigger a variety of transcriptional responses depending on conditions. This implies the presence of additional mechanisms, which modulate the activity of the corresponding TFs. Therefore, investigation of these mechanisms is relevant. Using ChIP-seq data on the binding of ABF1, BES1, MYC2 and ARR10 (which regulate the responses to abscisic acid, brassinosteroids, jasmonates and cytokinin, respectively), as well as the transcriptomes induced by these hormones, we were able to identify an enrichment of motifs, which are similar to the known binding sites for these TFs. All of them were associated with the increase of gene expression in response to hormones. Additionally, other motifs were enriched in the ChIP-seq peaks, which likely correspond to the binding sites of the potential partner TFs. For ABF1 and MYC2, functional composite elements were predicted, which likely are the binding sites for these TFs being a part of heterodimers with the partner TFs. It is noteworthy that the classic TF binding site for ARR10 was not the most enriched, at the same time, ARR10 had the largest number of the motifs for the potential partner TFs compared to other TFs. This may indicate that the functionality of TFs ABF1, BES1 and MYC2 is regulated primarily by the nucleotide context of their binding site, while partner TFs play a fundamental role in the activity of ARR10.

Conclusion: SFmotif pipeline has been developed for a comprehensive study of the mechanisms of regulation of gene activity. It was predicted that phytohormones can implement different strategies of transcriptional regulation with different contributions of partner transcription factors.

Funding: The study was supported by RSF project No. 20-14-00140.

Список литературы/References

1. Heinz S., Benner C., Spann N., Bertolino E., Lin Y.C., Laslo P., Cheng J.X., Murre C., Singh H., Glass C.K. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-89. doi 10.1016/j.molcel.2010.05.004
2. Levitsky V., Zemlyanskaya E., Oshchepkov D., Podkolodnaya O., Ignatieva E., Grosse I., Mironova V., Merkulova T. A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package. *Nucleic Acids Res.* 2019;47(21):e139. doi 10.1093/nar/gkz800
3. Yin L., Zander M., Huang S.C., Xie M., Song L., Saldierna Guzmán J.P., Hann E., Shanbhag B.K., Ng S., Jain S., Janssen B.J., Clark N.M., Walley J.W., Beddoe T., Bar-Joseph Z., Lewsey M.G., Ecker J.R. Transcription factor dynamics in cross-regulation of plant hormone signaling pathways. *bioRxiv*. 2023. doi 10.1101/2023.03.07.531630
4. Xie M., Chen H., Huang L., O'Neil R.C., Shokhirev M.N., Ecker J.R. A B-ARR-mediated cytokinin transcriptional network directs hormone cross-regulation and shoot development. *Nat Commun.* 2018;9(1):1604. doi 10.1038/s41467-018-03921-6

doi 10.18699/bgrs2024-1.2-04

Вычислительный конвейер по распознаванию *de novo* сайтов связывания транскрипционных факторов в бактериальных геномах

Мухин А.М.[1, 2, 3]*, Ощепков Д.Ю.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

[3] *Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия*

* *mukhin@bionet.nsc.ru*

Ключевые слова: Поиск *de novo* мотивов, филогенетический футпринтинг, вычислительный конвейер, Python, Nextflow

Мотивация и цель: Аннотация бактериальных геномов и их конкретных регуляторных геномных последовательностей сайтами связывания транскрипционных факторов (ССТФ) является актуальной биологической задачей, поскольку работа клетки и производство определенных ферментов критическим образом зависит от существующих регуляторных геномных связей. Существующие программные решения содержат необходимые программные компоненты лишь частично, не позволяя комплексно и с минимальными трудозатратами выполнять массовый поиск ССТФ как во вновь секвенированных, так и недостаточно изученных бактериальных геномах. В данной работе реализован вычислительный конвейер, состоящий из следующих необходимых этапов для полноценной аннотации бактериальных геномов с помощью известных подходов *de novo* поиска ССТФ: (1) аннотирование генома оперонной структурой, необходимое для дальнейшего точного определения регуляторных/промоторных областей, (2) поиск *de novo* мотивов в промоторах целевого генома, (3) функциональной аннотации вновь выявленных ССТФ. Поиск *de novo* мотивов в промоторах целевого генома может осуществляться альтернативно с помощью двух подходов: либо в полной выборке промоторов целевого организма, либо на основании подхода филогенетического футпринтинга, осуществляя поиск ССТФ в наборе промоторов ортологичных генов из одной таксономической группы с целевым организмом. В последнем случае необходимый список этапов аннотации должен включать также (4) инструмент для поиска ортологичных генов заданного таксономического уровня и (5) базы данных с последовательностями и аннотациями известных бактериальных геномов, что позволит автоматически осуществлять все необходимые операции по формированию требуемых выборок промоторов. Такое комплексное решение подразумевает также наличие всех необходимых программных модулей, осуществляющих формирование необходимых выборок, операции по конвертации форматов и перенаправлению данных и сохранения необходимых данных в служебных БД. Такой комплексный подход позволит сократить до минимума затраты ресурсов на промежуточные, но требующие квалификации в программировании для персонала с другой (микробиологической) квалификацией.

Методы и алгоритмы: Оба альтернативных конвейера поиска *de novo* мотивов - по полной выборке промоторов целевого организма и с помощью подхода филогенетического футпринтинга были реализованы с помощью следующих инструментов: язык программирования Python, база данных SQLite и PostgreSQL, системы конвейеризации NextFlow.
Результаты: Реализован конвейер частично с использованием платформы nextflow и набора скриптов на языках программирования Python и bash. На этапе по определению оперонной структуры используется альтернативно либо БД DOOR 2.0 [1] для известных организмов, либо веб-сервис Operon Mapper [2] для вновь секвенированных геномов. Для поиска *de novo* мотивов по полной выборке промоторов целевого организма используется инструмент BoBro2 [3], подход филогенетического футпринтинга реализован внутри программного модуля MP3, объединяющего значительное количество известных подходов для выявления ССТФ de novo в бактериальных геномах [4]. Также для построения выборок промоторов ортологичных генов используется инструмент для поиска ортологов внутри заданной таксономической группы на основе их белковых последовательностей GOST [5]. Необходимый этап функциональной аннотации вновь выявленных ССТФ осуществляется с помощью приложения Tomtom программного пакета MEME suite [6] путем сравнения выявленных мотивов с известными мотивами из соответствующих БД. Для повышения скорости расчетов была развернута база данных SQLite использующаяся для индексации координат генов, оперонов, данных по таксономической принадлежности геномов и пр.
Выводы: Создан вычислительный конвейер по распознаванию *de novo* сайтов связывания транскрипционных факторов в бактериальных геномах на основе двух подходов: по полной выборке промоторов целевого организма, либо на основании подхода филогенетического футпринтинга. Конвейер включает необходимый набор инструментов для осуществления всех необходимых промежуточных вычислений, который можно использовать как в локальной среде, так на кластере. В дальнейшем планируется интеграция дополнительного модуля по определению оперонной структуры с использованием алгоритма машинного или глубокого обучения; включение в программный комплекс других известных подходов для поиска ССТФ de novo, интеграцию дополнительных средств отображения, обработки и ранжирования получаемых данных; расширение интегрированной базы данных для сохранения результатов расчетов, требующих значительных вычислительных мощностей; замена СУБД на PostgreSQL.
Финансирование: Исследование поддержано в рамках Программы Курчатовского геномного центра ИЦиГ СОРАН (№ 075-15-2019-1662).

A computational pipeline for *de novo* recognition of transcription factor binding sites in bacterial genomes

Mukhin A.M. [1, 2, 3]*, Oshchepkov D.Y.[1, 2]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center ICiG SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk National Research State University, Novosibirsk, Russia*

* *mukhin@bionet.nsc.ru*

Ключевые слова: de novo motifs, phylogenetic footprinting, computational pipeline, Python, Nextflow

Motivation and aim: Annotation of bacterial genomes and their specific regulatory genomic sequences by transcription factor binding sites (TFBS) is an urgent biological task, since cell function and production of certain enzymes critically depend on existing regulatory genomic connections. Existing software solutions contain the necessary software components only partially, not allowing a comprehensive and labor-intensive mass search for TFBS in both newly sequenced and insufficiently studied bacterial genomes. In this work, we have implemented a computational pipeline consisting of the following necessary steps for full annotation of bacterial genomes using known *de novo* search approaches for TFBS: (1) annotation of the genome with operon structure necessary to further pinpoint regulatory/promoter regions, (2) *de novo* search for motifs in the promoters of the target genome, and (3) functional annotation of newly identified TFBS. The *de novo* motifs search in the promoters of the target genome can be performed alternatively using two approaches: either in the full set of promoters of the target organism, or based on a phylogenetic footprinting approach, searching for TFBS in a set of promoters of orthologous genes from the same taxonomic group as the target organism. In the latter case, the required list of annotation steps should also include (4) a tool to search for orthologous genes of a given taxonomic level and (5) databases with sequences and annotations of known bacterial genomes, allowing to automatically perform all necessary operations to generate the required promoter samples. Such a complex solution also implies the presence of all necessary program modules, which perform the formation of the required samples, operations on format conversion and data redirection and saving the required data in the service database. Such an integrated approach will minimize the resources spent on intermediate, but requiring programming skills for personnel with other (microbiological) qualifications.
Methods and algorithms: Both alternative pipelines for *de novo* motif search - by full sampling of target organism promoters and by phylogenetic footprinting approach - were implemented using the following tools: Python programming language, SQLite and PostgreSQL database, NextFlow pipelining system.
Results: The pipeline is implemented partially using the nextflow platform and a set of scripts in Python and bash programming languages. The operon structure determination step alternatively uses either the DOOR 2.0 database [1] for known organisms or the web service Operon Mapper [2] for newly sequenced genomes. The BoBro2 tool [3] is used to search for de novo motifs from a full sample of promoters of the target organism; the phylogenetic footprinting approach is implemented within the MP3 software module, which integrates a significant number of known approaches for detecting *de novo* TFBS in bacterial genomes [4]. Also, a tool GOST for searching orthologous genes within a given taxonomic group based on their protein sequences is used to construct orthologous gene promoter samples [5]. The necessary step of functional annotation of newly identified TFBS is performed using the Tomtom application of the MEME suite software [6] by comparing the identified motifs with known motifs from the corresponding databases. To increase the speed of calculations, we deployed a SQLite database used for indexing the coordinates of genes, operons, data on the taxonomic affiliation of genomes, etc.
Conclusions: A computational pipeline for de novo recognition of transcription factor binding sites in bacterial genomes has been developed based on two approaches: full sampling of target organism promoters or phylogenetic footprinting. The pipeline includes the necessary set of tools to perform all necessary intermediate calculations, which can be used both in a local environment and on a cluster. The future plans include

the following steps: (a) to integrate an additional module for operon structure determination using a machine learning or deep learning algorithm; (b) to include other known approaches for *de novo* search of TFBS into the program complex; (c) to integrate additional tools for displaying, processing, and ranking the obtained data; (d) to expand the integrated database for saving the results of calculations that require significant computing power; (e) to replace the DBMS with PostgreSQL.

Funding: The research was supported under the Kurchatov Genomic Center Program of ICG SB RAN (No. 075-15-2019-1662).

Список литературы/References

1. Mao X. et al. DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Res*. 2014;42(D1):D654-D659
2. Taboada B. et al. Operon-mapper: A web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*. 2018;34(23):4118-4120
3. Ma Q. et al. An integrated toolkit for accurate prediction and analysis of cis-regulatory motifs at a genome scale. *Bioinformatics*. 2013;29(18):2261-2268
4. Liu B. et al. An integrative and applicable phylogenetic footprinting framework for cis-regulatory motifs identification in prokaryotic genomes. *BMC Genomics.* 2016;17:578
5. Li G. et al. Integration of sequence-similarity and functional association information can overcome intrinsic problems in orthology mapping across bacterial genomes. *Nucleic Acids Res*. 2011;39(22):e150
6. Bailey T.L. et al. The MEME Suite. *Nucleic Acids Res*. 2015;43(W1):W39-W49

doi 10.18699/bgrs2024-1.2-05

Предсказание активности и рациональный дизайн регуляторных районов генома и транскриптома эукариот методами машинного обучения

Пензар Д.Д.[1, 2]*, Ногина Д.Д.[1], Носкова Е.О.[1], Власов А.В.[1], Зинкевич А.О.[1], Кулаковский И.В.[3]

[1] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

[2] *Институт искусственного интеллекта AIRI, Москва, Россия*

[3] *Институт белка РАН, Пущино, Россия*

* *dmitrypenzar1996@gmail.com*

Ключевые слова: регуляторная геномика; глубокое машинное обучение; сверточные сети; регуляторные варианты; рациональный дизайн

Мотивация и цель: Исследования по расшифровке структуры и вычислительному предсказанию активности регуляторных районов генома эукариот сегодня развиваются особенно бурно благодаря синергии высокопроизводительных экспериментальных методов и машинного обучения. Накапливается все больше данных массовых параллельных экспериментов с репортерами, которые позволяют оценить активность сотен тысяч и даже миллионов регуляторных элементов. Неотъемлемым инструментом для работы с такими данными является машинное обучение, в особенности глубокие нейронные сети. Ранее мы предложили адаптацию архитектуры EfficientNetV2 [1], используемой для классификации изображений, к анализу нуклеотидных последовательностей. Предложенная нами модель LegNet [2] с большим отрывом победила конкурирующие решения на основе рекуррентных сетей и сетей на основе механизма внимания в задаче предсказания экспрессии репортерного белка в клетках дрожжей по последовательности промотора.

Сегодня нам удалось адаптировать модель к широкому спектру геномных задач: от предсказания активности регуляторных последовательностей в различных типах клеток человека до оценки эффекта однонуклеотидных замен в них и рационального дизайна регуляторных последовательностей с заданными свойствами.

Методы и алгоритмы: LegNet – полностью конволюционная нейронная сеть для работы с нуклеотидными последовательностями [2], реализованная на фреймворке pytorch [3]. В ходе обучения нейронной сети использовался метод one-cycle learning rate schedule [4]. Для обучения нейросети использовали данные из работ [5–7], для каждой из задач архитектура адаптировалась исходя из размера обучающей выборки, предсказываемой переменной и размера последовательностей. Для валидации использовали данные из работ [5–10].

Результаты: LegNet показывает SOTA (state-of-the-art) результаты на задаче предсказания экспрессии по регуляторной последовательности для промоторов дрожжей, регуляторных элементов дрозофилы, энхансеров человека. Для регуляторных районов генома человека достигнутое качество модели превосходит Enformer [11], предобученной на большом количестве экспериментов из базы

ENCODE. Модели, обученные для регуляторных районов генома человека, способны предсказать события аллель-специфичности по данным ADASTRA [8] и UDACHA [9]. Также мы впервые предложили и успешно протестировали генеративную модель на основе диффузии для генерации промоторов с заданным уровнем экспрессии, не ограничиваясь работой с категориальными переменными [12, 13]. Наконец, мы адаптировали LegNet для работы на коротких последовательностях и, используя данные высокопроизводительного SELEX, показали применимость в задаче предсказания эффектов однонуклеотидных вариантов, измеренных в эксперименте SNP-SELEX [10].
Выводы: Нами была разработана новая архитектура сверточной нейронной сети LegNet и методика ее обучения, позволяющие достигнуть SOTA результатов на большом спектре задач регуляторной геномики. Предложенный метод применим для разных длин нуклеотидных последовательностей и разных организмов для решения задач предсказания активности регуляторных районов и для дизайна последовательностей с заданными свойствами.
Финансирование: Исследование было поддержано грантом РНФ 20-74-10075.

Machine learning for rational design and reliable prediction of activity of gene regulatory regions

Penzar D.D.[1, 2]*, Nogina D.D.[1], Noskova E.O.[1], Vlasov A.V.[1], Zinkevich A.O.[1], Kulakovskiy I.V.[3]

[1] *Lomonosov Moscow State University, Moscow, Russia*
[2] *AIRI, Moscow, Russia*
[3] *Institute of Protein Research, RAS, Pushchino, Russia*
* *dmitrypenzar1996@gmail.com*

Key words: regulatory genomics; deep learning; CNN rational design; regulatory variants

Motivation and Aim: Synergy of high-throughput experimental methods and machine learning has substantially driven the recent advances in deciphering the structure and predicting the activity of eukaryotic gene regulatory regions. The volume of available data obtained from massively parallel reporter assays is growing fast, and the already available dataset include activity estimates for hundreds of thousands or even tens of millions of diverse regulatory elements. Machine learning, in particular, deep neural networks became an essential tool to properly utilize such large-scale data for modeling gene regulatory regions. In previous studies, we had shown how to adapt the EfficientNetV2 architecture [1], originally used for image classification, to be effectively applicable for the analysis of nucleotide sequences. Our LegNet model [2] outperformed multiple competing solutions, including those based on recurrent networks and attention-based networks, in solving the problem of predicting reporter protein expression in yeast cells solely from the promoter sequence.
Here we show that LegNet is applicable to a wide range of genomic problems from predicting the activity of regulatory sequences in various human cell types to assessing the effect of single nucleotide variants and rational design of regulatory sequences with desired properties.

Methods and Algorithms: LegNet is a fully convolutional neural network optimized to handle nucleotide sequences [2] and implemented with PyTorch [3] using the one-cycle learning rate schedule for training [4]. Data from [5–7] was used for model training for different tasks, while the architecture was modified to account for varying sizes of the training set, the target variable, and the lengths of sequences. Data from [5–10] was used as independent test data.
Results: LegNet demonstrates state-of-the-art (SOTA) results in predicting expression from the regulatory sequence for yeast promoters, estimating the activity of Drosophila regulatory elements, and human enhancers. For human gene regulatory regions, the achieved model performance surpasses Enformer [11], which was pre-trained on a large number of experiments from the ENCODE database. Further, models trained for regulatory regions of the human genome are capable of predicting allele-specific events based on data from ADASTRA [8] and UDACHA [9].
Also, we have proposed and tested a generative diffusion-based model for generating promoters with a desired level of expression, the first of its kind not limited by handling categorical variables only [12, 13]. Finally, we have adapted LegNet to shorter sequences and, with high-throughput SELEX data, demonstrated its capability in recognizing regulatory single nucleotide variants dected by the SNP-SELEX experiment [10].
Conclusion: We have developed a new convolutional neural network architecture, LegNet, and a training scheme allowing to achieve SOTA results in a wide range of machine learning applications in eukaryotic regulatory genomics. The proposed method is applicable to sequences of different lengths and for different organisms, enabling the prediction of regulatory region activity and the design of sequences with desired properties.
Funding: This study was supported by the Russian Science Foundation grant (20-74-10075 to I.V.K.)

Список литературы/References

1. Tan M., Le Q.V. EfficientNetV2: Smaller Models and Faster Training. [Preprint]. *Arxiv*. 2021. doi 10.48550/arXiv.2104.00298
2. Penzar D., Nogina D., Noskova E., Zinkevich A., Meshcheryakov G., Lando A., Rafi A.M., de Boer C., Kulakovskiy I.V. LegNet: a best-in-class deep learning model for short DNA regulatory regions. *Bioinformatics*. 2023;39(8):btad457. doi 10.1093/bioinformatics/btad457
3. Paszke A., Gross S. et al. Automatic differentiation in PyTorch. *Open Rev*. 2017. https://openreview.net/pdf/25b8eee6c373d48b84e5e9c6e10e7cbbbce4ac73.pdf?ref=blog.premai.io
4. Smith L.N., Topin N. Super-convergence: very fast training of neural networks using large learning rates. [Preprint]. *Arxiv*. 2017. doi 10.48550/arXiv.1708.07120
5. Rafi A.M., Nogina D., Penzar D. et al. Evaluation and optimization of sequence-based gene regulatory deep learning models. *bioRxiv*. 2023. doi 10.1101/2023.04.26.538471
6. Agarwal V., Inoue F., Schubach M., Martin B.K., Dash P.M., Zhang Z., Sohota A., Noble W.S., Yardimci G.G., Kircher M., Shendure J., Ahituv N. Massively parallel characterization of transcriptional regulatory elements in three diverse human cell types. *bioRxiv*. 2023. doi 10.1101/2023.03.05.531189
7. de Almeida B.P., Reiter F., Pagani M. et al. DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. *Nat Gen.* 2022;54:613-624. doi 10.1038/s41588-022-01048-5
8. Abramov S., Boytsov A., Bykova D. et al. Landscape of allele-specific transcription factor binding in the human genome. *Nat Commun.* 2021;12:2751. doi 10.1038/s41467-021-23007-0
9. Buyan A. et al. *S*tatistical framework for calling allelic imbalance in high-throughput sequencing data. b*ioRxiv*. 2023. doi 10.1101/2023.11.07.565968
10. Yan J., Qiu Y. et al. Systematic analysis of binding of transcription factors to noncoding variants. *Nature*. 2021;591(7848):147-151. doi 10.1038/s41586-021-03211-0

doi 10.18699/bgrs2024-1.2-06

Применение эволюционных аугментаций для работы с геномными последовательностями

Пензин Н.[1], Пензар Д.[1, 2]*

[1] *Московский государственный университет им. М.В. Ломоносова, Москва Россия*

[2] *Институт белка РАН, Пущино, Россия*

* *dmitrypenzar1996@gmail.com*

Ключевые слова: аугментации; EvoAug; LegNet

Мотивация и цель: В прошлом году американскими исследователями из Центра количественной биологии Симонса (Simons Center for Quantitative Biology) был опубликован фреймворк EvoAug на PyTorch с открытым исходным кодом [1]. В биологии частой проблемой является недостаточность данных для обучения глубокой нейронной сети (DNN), вследствии чего нейронная сеть склонна переобучаться и показывать плохую способность к генерализации на новых данных. Авторы показывают, что обучение нейронных сетей с помощью фреймворка EvoAug, применяющего в ходе обучения набор аугментаций, основанных на псевдоэволюционных процессах, приводит к улучшению генерализации моделей и повышает эффективность стандартных методов апостериорного сравнения, например, интерпретируемости фильтров и анализа параметров, в решении важных задач регуляторной геномики.

Целью данной работы является проверка эффективности и апробация фреймворка для различных задач регуляторной геномики.

Методы и алгоритмы:

- Basset – конволюционная нейронная сеть, используемая в качестве многозадачной бинарной классификации сайтов доступности хроматина в 161 типе клеток [2].
- DeepSTARR – конволюционная нейронная сеть, представляющая собой многозадачную количественную регрессию для прогнозирования активности энхансера на основе данных секвенирования самотранскрибирующейся активной регуляторной области (STARR-seq) [3].
- LegNet – конволюционная нейронная сеть для предсказания экспрессии генов из массового параллельного репортерного анализа [4].
- Данные массового параллельного репортерного анализа для трех основных типов клеток ENCODE: гепатоциты человека (HepG2), лимфобласты (K562) и индуцированные плюрипотентные стволовые клетки (WTC11) [5].

Результаты: Авторы данного фреймворка показывают улучшение генерализации моделей Basset и DeepSTARR с использованием набора аугментаций, предоставляемых их алгоритмом. Нами были воспроизведены результаты, представленные в статье, и сделаны выводы о действительном улучшении качества предсказаний для задач классификации сайтов доступности хроматина и прогнозирования активности энхансера.

Рассматриваемый фреймворк позволил также повысить качество модели, направленной на предсказание эффективности трансляции у прокариот. Также EvoAug был испробован для задачи предсказания экспрессии эукариотических

генов для человека из массового параллельного репортерного анализа и для дрожжей из гигантского параллельного репортерного анализа. Были подобраны оптимальные гиперпараметры для каждого типа аугментации, поддерживаемого алгоритмом. Обучение модели с набором аугментаций, предоставляемых фреймворком, не позволило улучшить результаты исходной нейронной сети для обеих задач.
Выводы: Для части задач регуляторной геномики фреймворк EvoAug действительно позволяет получать модель с лучшим качеством, нежели исходные. Однако для части датасетов может вообще не приводить к улучшению качества предсказания, а дообучение может не позволить избавиться от искажений в данных, вносимых аугментациями.

Application of evolutionary augmentations to work with genomic sequences

Penzin N.[1], Penzar D.[1, 2]*
[1] *Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow, Russia*
[2] *Institute of Protein Research, RAS, Pushchino, Russia*
* *dmitrypenzar1996@gmail.com*

Key words: augmentation; EvoAug; LegNet

Motivation and Aim: Last year the EvoAug, an open-source PyTorch package, was published by American researchers from the Simons Center for Quantitative Biology [1]. A common biology problem is insufficient data to train a deep neural network (DNN), causing the neural network to overtrain and exhibit poor generalization ability on new data. The authors of the original paper show that training neural networks using the EvoAug framework, that provides a suite of evolution-inspired data augmentations, leads to better generalization performance and improves efficacy with standard post hoc explanation methods, including filter interpretability and attribution analysis, across prominent regulatory genomics prediction tasks for well-established DNNs.
The purpose of this work is to test the effectiveness and test the framework for various tasks of regulatory genomics.
Methods and Algorithms:

- Basset is a convolutional neural network used as a multi-task binary classification of chromatin accessibility sites in 161 cell types [2].
- DeepSTARR is a convolutional neural network that is a multi-task quantitative regression for predicting enhancer activity based on self-transcribing active regulatory region sequencing (STARR-seq) data [3].
- LegNet is a convolutional neural network for predicting gene expression from yeast gigantic parallel reporter assays [4].
- Massive parallel reporter assay data for three major ENCODE cell types: human hepatocytes (HepG2), lymphoblasts (K562) and induced pluripotent stem cells (WTC11) [5].

Results: The authors of this framework show improved generalization of the Basset and DeepSTARR models using the set of augmentations provided by their algorithm. We reproduced the results presented in the article, additionally tested them against more

strict baseline methods and drew conclusions about a real improvement in the quality of predictions for the tasks of classifying chromatin accessibility sites and predicting enhancer activity. Next we tested EvoAug on a task of prediction translation efficiency in prokaryotes that resulted in improved model performance.

After that we decided to apply EvoAug for eukaryotic gene expression prediction using data from a massively parallel reporter assay. Optimal hyperparameters were selected for each type of augmentation provided by the algorithm. Training the model with a set of augmentations provided by the framework did not improve the results of the original neural network.

As the last step we applied the framework for the yeast GPRA promoter dataset and showed that using the framework for this task results in suboptimal results.

Conclusion: For some regulatory genomic problems the EvoAug framework actually allows one to obtain a model with better quality than the original ones. However, for some datasets it may not lead to an improvement in the quality of prediction at all or even result in suboptimal results. Moreover, further model fine-tuning suggested by authors may not allow it to get rid of bias in the data introduced by augmentations and in general doesn't result in performance gains.

Список литературы/References

1. Lee N.K., Tang Z., Toneyan S. et al. EvoAug: improving generalization and interpretability of genomic deep neural networks with evolution-inspired data augmentations. *Genome Biol.* 2023;24(1):105. doi 10.1186/s13059-023-02941-w
2. Penzar D., Nogina D., Noskova E., Zinkevich A., Meshcheryakov G., Lando A., Rafi A.M., de Boer C., Kulakovskiy I.V. LegNet: a best-in-class deep learning model for short DNA regulatory regions. *Bioinformatics*. 2023;39(8):btad457. doi 10.1093/bioinformatics/btad457
3. Kelley D.R., Snoek J., Rinn J.L. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res*. 2016;26(7):990-999. doi 10.1101/gr.200535.115
4. de Almeida B.P., Reiter F., Pagani M., Stark A. DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. *Nat Genet*. 2022;54(5):613-624. doi 10.1038/s41588-022-01048-5
5. Agarwal V., Inoue F., Schubach M. et al. Massively parallel characterization of transcriptional regulatory elements in three diverse human cell types. *bioRxiv*. 2023. doi 10.1101/2023.03.05.531189

doi 10.18699/bgrs2024-1.2-07

Разработка компьютерных инструментов для поиска генов-мишеней микроРНК

Радочин П.А. [1], Воронина А.А. [2], Орёл Д.Ю. [2], Смирнова О.А. [1], Голубчиков Д.О.[1], Мак Д.В. [1], Воронова С.С. [1], Волков И.А.[1], Савина Е.А. [1, 3], Орлов Ю.Л.[1, 4]*

[1] *Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия*

[2] *Санкт-Петербургский политехнический университет им. Петра Великого, Санкт-Петербург, Россия*

[3] *Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

[4] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

** orlov@bionet.nsc.ru*

Ключевые слова: биоинформатика; микроРНК; базы данных; генные сети; комплексные заболевания; генные онтологии; образование

Мотивация и цель: Исследование молекулярных механизмов комплексных заболеваний является важной задачей биомедицины, требующей интеграции данных, развития новых биоинформационных инструментов [1]. микроРНК – это класс некодирующих молекул РНК, которые регулируют трансляцию и деградацию своих МРНК-мишеней посредством взаимодействия, что делает изучение их функций при заболеваниях необходимым для выяснения их патогенетических механизмов [2]. миРНК играют важную роль в регуляции экспрессии генов. Появляются все новые литературные данные о роли микроРНК в прогрессии опухолей, развитии ментальных расстройств, заболеваниях печени, хронических заболеваниях [1].

Представлен научный проект исследования генов-мишеней микроРНК в рамках Цифровой кафедры Сеченовского Университета в 2023/2024 учебном году [3]. Освоение использования существующих онлайн-программ для компьютерного поиска генов-мишеней микроРНК в геноме и разработка собственных программ для предсказания участков связывания микроРНК в генах.

Решаемая проблема в рамках представленного проекта – обучение по проблемам использования инструментов биоинформатики для анализа структуры РНК и микроРНК, механизмам связывания РНК с транскриптами генов человека.

Методы и алгоритмы: Собраны имеющиеся программные инструменты по поиску генов-мишеней микроРНК, подготовлен конвейер обработки данных. Использовались обзоры (Список программ и баз – https://www.tamirna.com/micrornas-web-based-tools/), базы данных MIRDB (https://mirdb.org/), TargetScan (https://www.targetscan.org/vert_80/), MiRTarBase (https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2022/php/index.php).

микроРНК – это РНК, состоящая из ~22 нуклеотидов, обработанных из шпилечных структур РНК. микроРНК слишком короткие, что создает технические трудности их поиска. Определение генов мишеней по частичной комплементарности также осложняется возможностью несовпадения нуклеотидов [4].

Составление списков генов и микроРНК, связанных с заболеваниями, выполнялось на основе запросов к базам данных OMIM (Online Mendelian Inheritance in Man), GeneCards, MalaCards. Анализ генных онтологий для списка

генов был выполнен с помощью ресурса PANTHER (Protein ANalysis THrough Evolutionary Relationships) (http://pantherdb.org/) и DAVID. Для реконструкции сети использовались инструменты STRING-DB, приложения Cytoscape. С помощью онлайн-инструментов биоинформатики был проанализирован актуальный массив данных.

Результаты: TargetScan предсказывает биологические мишени микроРНК путем поиска консервативных 8-, 7- и 6-мерных сайтов, которые соответствуют начальной области каждой микроРНК. В качестве вариантов также предлагаются предсказания только со слабоконсервативными сайтами и предсказания с несохраняемыми микроРНК [5]. Также идентифицированы сайты с несоответствиями в начальной области, которые компенсируются консервативным 3'-спариванием. У млекопитающих прогнозы ранжируются на основе прогнозируемой эффективности нацеливания, рассчитанной с использованием биохимической модели репрессии, опосредованной микроРНК. Выявлены общие черты, связанные со связыванием микроРНК и снижением регуляции мишеней, которые были использованы для прогнозирования мишеней для микроРНК с помощью методов машинного обучения.

Для поиска генов-мишеней терапии необходима реконструкция генной сети – комплекса взаимодействующих макромолекул, включая миРНК, функционально связанных с заболеванием. Дальнейший анализ предполагает кластеризацию генов в сети, выявление ключевых генов, обладающих наибольшим числом контактов в сети, нахождение категорий генных онтологий.

Представлен конвейер использования программных инструментов для поиска генов-мишеней микроРНК для медицины. Возможно дальнейшее использование таких программ в качестве учебных ресурсов по биоинформатике для медиков.

Работы были представлены на студенческой конференции в Сеченовском Университете, подготовлены студенческие дипломы по теме анализа микроРНК с помощью онлайн-инструментов биоинформатики.

Выводы: Представлена методика биоинформационного исследования. Использование микроРНК как регуляторных макромолекул представляет перспективную научную тему в биотехнологии и медицине. Анализ предсказанных генов-мишеней позволит найти новые таргеты (мишени) для терапии. Определены гены мишени микроРНК, известных из научной литературы, для ряда заболеваний. Практическое использование найденных микроРНК как маркеров для диагностики заболеваний, в практических проектах по высокотехнологической медицине остается сложной задачей, требующей дальнейшей интеграции ресурсов.

Благодарности: Исследование поддержано проектом Цифровая кафедра Сеченовского Университета.

Development of computer tools to search for microRNA targets

Radochin P.A.[1], Voronina A.A.[2], Orel D.Y.[2], Smirnova O.A.[1], Golubchikov D.O.[1], Mak D.V.[1], Voronova S.S.[1], Volkov I.A.[1], Savina E.A.[1, 3], Orlov Y.L.[1, 4]*

[1] *Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

[2] *St. Petersburg Polytechnical University, St. Petersburg, Russia*

[3] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

[4] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *orlov@bionet.nsc.ru*

Key words: bioinformatics; microRNAs; databases; gene networks; complex diseases; gene ontologies; education

Motivation and Aim: The study of the molecular mechanisms of complex diseases is an important task of biomedicine, requiring data integration and the development of new bioinformatic tools [1]. microRNAs are a class of non-coding RNA molecules that regulate the translation and degradation of their MRNA targets through interaction, which makes the study of their functions in diseases necessary to elucidate their pathogenetic mechanisms [2]. miRNAs play an important role in regulating gene expression. New literature data are emerging on the role of microRNAs in tumor progression, the development of mental disorders, liver diseases, and chronic diseases [1].

A scientific project for the study of microRNA target genes within the Digital Department of Sechenov University in the 2023–2024 academic year is presented [3]. Mastering the use of existing online programs for computer search of microRNA target genes in the genome, and developing proprietary programs for predicting microRNA binding sites in genes.

The problem being solved within the framework of the presented project is training on the problems of using bioinformatics tools to analyze the structure of RNA and microRNA, the mechanisms of binding RNA to human gene transcripts.

Methods and Algorithms: The available software tools for searching for microRNA target genes have been collected, and a data processing pipeline has been prepared.

Reviews were used (List of programs and databases – https://www.tamirna.com/micrornas-web-based-tools/), MIRDB databases (https://mirdb.org/), TargetScan (https://www.targetscan.org/vert_80/), MiRTarBase (https://mirtarbase.cuhk.edu.cn /~miRTarBase/miRTarBase_2022/php/index.php),

microRNA is an RNA consisting of ~22 nucleotides processed from the hairpin structures of RNA. microRNAs are too short, which creates technical difficulties in finding them. The determination of target genes by partial complementarity is also complicated by the possibility of nucleotide mismatch [4].

Lists of genes and microRNAs associated with diseases were compiled based on queries to the OMIM (Online Mendelian Inheritance in Man), GeneCards, and MalaCards databases. The analysis of gene ontologies for the list of genes was performed using the PANTHER (Protein ANalysis THrough Evolutionary Relationships) resource (http://pantherdb.org/) and DAVID. STRING-DB tools and Cytoscape applications were used to reconstruct the network. An up-to-date data set was analyzed using online bioinformatics tools.

Results: We compared work of the tools for miRNA prediction. TargetScan predicts biological microRNA targets by searching for conserved 8-, 7-, and 6-dimensional sites that correspond to the initial region of each microRNA. Predictions with only weakly conservative sites and predictions with unsaved microRNAs are also offered as options [5]. Sites with inconsistencies in the initial region have also been identified, which are compensated by conservative 3'-pairing. In mammals, forecasts are ranked based on the predicted targeting efficiency calculated using a biochemical model of microRNA-mediated repression, common features associated with microRNA binding and decreased regulation of targets were identified, which were used to predict targets for microRNAs using machine learning methods.

To search for target genes for therapy, it is necessary to reconstruct the gene network – a complex of interacting macromolecules, including miRNAs, functionally related to the disease. Further analysis involves clustering genes in the network, identifying key genes with the largest number of contacts in the network, and finding categories of gene ontologies.

A pipeline of using software tools to search for microRNA target genes for medicine is presented. It is possible to further use such programs as educational resources on bioinformatics for physicians.

The works were presented at a student conference at Sechenov University, and student diplomas were prepared on the topic of microRNA analysis using online bioinformatics tools.

Conclusion: The technique of bioinformatics research is presented. The use of microRNAs as regulatory macromolecules is a promising scientific topic in biotechnology and medicine. The analysis of predicted target genes will allow us to find new targets for therapy. microRNA target genes known from scientific literature have been identified for a number of diseases. The practical use of the found microRNAs as markers for the diagnosis of diseases in practical projects on high-tech medicine remains a difficult task requiring further integration of resources.

Acknowledgements: The study was supported by the Digital Chair project of Sechenov University.

Список литературы/References

1. Beylerli O., Ilyasova T., Shi H., Sufianov A. MicroRNAs in meningiomas: Potential biomarkers and therapeutic targets. *Noncoding RNA Res.* 2024;9(3):641-648. doi 10.1016/j.ncrna.2024.02.011
2. McGeary S.E., Lin K.S., Shi C.Y., et al. The biochemical basis of microRNA targeting efficacy. *Science*. 2019;366(6472):eaav1741. doi 10.1126/science.aav1741
3. Turkina V.A., Orlova N.G., Orlov Y.L. Biophysics education section and computational training discussion at VII Congress of Russian Biophysicists. *Biophys Rev.* 2023;15:807-809. doi 10.1007/s12551-023-01147-5
4. Liu W., Wang X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol* 2019;20:18. doi 10.1186/s13059-019-1629-z
5. Peterson S.M., Thompson J.A., Ufkin M.L., Sathyanarayana P., Liaw L., Congdon C.B. Common features of microRNA target prediction tools. *Front. Genet.* 2014;5:23. doi 10.3389/fgene.2014.00023

doi 10.18699/bgrs2024-1.2-08

Enformer deep learning model for predicting mutation effects on gene transcriptional activity

Belokopytova P.[1, 2]*, Kuratov Y.[3], Fishman V. [1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *AIRI, Moscow, Russia*

* *belokopytova@bionet.nsc.ru*

Key words: deep learning; gene expression; enhancer–promoter interactions

Motivation and Aim: Investigation of mechanisms underlying the regulation of gene expression is one of the most relevant areas in modern biology. Genomic variants in the coding genomic regions can lead to the synthesis of defective proteins, which can be the cause of various diseases. However, there may also be mutations in non-coding genomic regions resulting in changes in transcriptional activity leading to pathology phenotype. Thus, the task of finding a connection between genomic variants and transcription activity of genes becomes especially actual. Currently, there are no convenient and accessible tools for evaluating the effects of mutations in non-coding genomic regions. The majority of current methods focus on the analysis of exomic variants, but due to the decreasing cost of sequencing, there is a gradual shift towards genome-wide analysis. This is making the study of single nucleotide variants and small structural rearrangements in non-coding regions of the genome increasingly accessible. This in turn requires the development of methods for their interpretation. In recent years, deep learning algorithms have been actively used in genomics, including the prediction of gene transcriptional activity.

Methods and Algorithms: We chose the state-of-the-art deep learning model Enformer [1], which was developed to predict gene transcriptional activity by DNA sequence. This model is based on a relatively new neural network architecture – Transformer. This neural network architecture makes it possible to integrate information about genomic interactions over a long distance (up to 200 kb bp). The attention layers of the neural network and *in silico* mutagenesis make it possible to determine the contribution of each nucleotide of the DNA sequence to the prediction. It has been shown that sequences with the greatest predictive weight are regulatory or insulatory nucleotide sequences. The ability to vary different genetic variants makes it possible to use the Enformer model to study quantitative trait loci (eQTL) without having to study hundreds or thousands of individual gene expression profiles. It is therefore promising to use the Enformer model as the basis for a tool that predicts changes in the transcriptional activity of genes associated with different genomic variants.

Results: We aimed to check how well the Enformer model deals with predicting changes of gene transcriptional activity on real experimental examples. In recent papers [2–4] has been shown that the Enformer model often incorrectly predicts the direction of transcriptional changes for individual genomes. It is worth noting that the level of gene transcription varies very slightly among healthy people and such changes of transcriptional activity are difficult to detect. There weren't any investigations of the Enformer model ability to predict a critical change in gene expression in the case of

single nucleotide or short genomic mutations and chromosomal rearrangements leading to pathology. We analyzed the literature and created a dataset that includes mutations in promoters and enhancers of genes leading to critical gene expression changes (at least two times).

Conclusion: We have shown that in the most investigated cases the Enformer model predicts right direction and level of transcriptional activity changes. This fact allows creating a tool for interpreting non-coding genomic variants that will expand the available information about genetic anomalies obtained by different genomic analyses. This may help medical genetics to put forward new hypotheses regarding the molecular mechanisms associated with the pathological phenotype.

Funding: Ministry of Education of Russia project № 075-15-2024-539.

References

1. Avsec Ž., Agarwal V., Visentin D. et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nat Methods*. 2021;18(10):1196-1203. doi 10.1038/s41592-021-01252-x
2. Huang C. et al. Personal transcriptome variation is poorly explained by current genomic deep learning models. *Nat Genet*. 2023;55(12):2056-2059
3. Sasse A. et al. Benchmarking of deep neural networks for predicting personal gene expression from DNA sequence highlights shortcomings. *Nat Genet*. 2023;5(12):2060-2064
4. Tang Z., Toneyan S., Koo P.K. Current approaches to genomic deep learning struggle to fully capture human genetic variation. *Nat Genet*. 2023;55(12):2021-2022

doi 10.18699/bgrs2024-1.2-09

Search for the transcription factors involved in the regulation of biofilm formation in *Escherichia coli* K-12

Bessonova T.A.[1, 2]*, Dakhnovets A.I.[3], Kuznetsova U.D.[4], Ozoline O.N.[1], Gelfand M.S.[3], Tutukina M.N.[1, 2, 3]

[1] *Institute of Cell Biophysics, RAS, Pushchino Scientific Center for Biological Research of RAS, Pushchino, Russia*

[2] *Institute for Information Transmission Problems, RAS (Kharkevich Institute), Moscow, Russia*

[3] *Skolkovo Institute of Science and Technology, Moscow, Russia*

[4] *Pirogov Russian National Research Medical University, Moscow, Russia*

* *tatianabessonova66@gmail.com*

Key words: biofilm; transcription factors; non-coding RNAs; *E. coli*; UxuR; cAMP-CRP; YjjM (CsqR)

Motivation and Aim: Bacterial biofilms are one of the main reasons for the worldwide expanding antibiotic resistance. It takes 1,000 times as much antibiotic to treat an infection if bacteria "decide" to form a biofilm. Lots of studies are being conducted for *Pseudomonas* and *Staphylococcus*. However, the regulation of biofilm formation in *Escherichia coli*, which can also be pathogenic, is still poorly understood. Uropathogenic *E. coli* (UPEC), forming extensive biofilms, leads to chronic urinary tract infections, while the biofilms formed by EPEC can be the reason for chronic colitis or gastroenteritis. The laboratory opportunistic *E. coli* K-12 MG1655 strain is also characterized by biofilm formation. A variety of candidate regulators of this process have been proposed but the master regulator is not yet known. The main aim of our research is to find the key factors promoting changes of the *E. coli* lifestyle, and to study the possible connection between biofilm formation and metabolic features.

Methods and Algorithms: RNA was extracted from the free-living *E. coli* cells and biofilms, and RNA-seq was performed in triplicates. After QC, reads were mapped on the reference genome using Bowtie2, and differential expression was analysed with FeatureCounts and DESeq2. Biofilm intensity was measured in 96 well plates after staining with 0.1 % crystal violet. Bacterial adhesion efficiency was tested on the CaCo-2 monolayer. Target gene expression was evaluated using qRT-PCR. Calculations were performed using $2^{-\Delta\Delta Ct}$.

Results: Based on the RNA-seq data, regulators of hexuronate metabolism, UxuR and YjjM, could serve as repressors of the *E. coli* motility and activators of biofilm formation. To form a biofilm, bacteria first need to attach to a surface. As such, we measured the adhesion ability of the wild type *E. coli* K-12 MG1655 and its *ΔuxuR, ΔexuR, ΔyjjM,* and *Δcrp* derivatives to the intestinal epithelial cells CaCo-2 (Fig. 1). cAMP-CRP was earlier shown to be involved in regulation of these processes, and ExuR is the UxuR paralog [1].

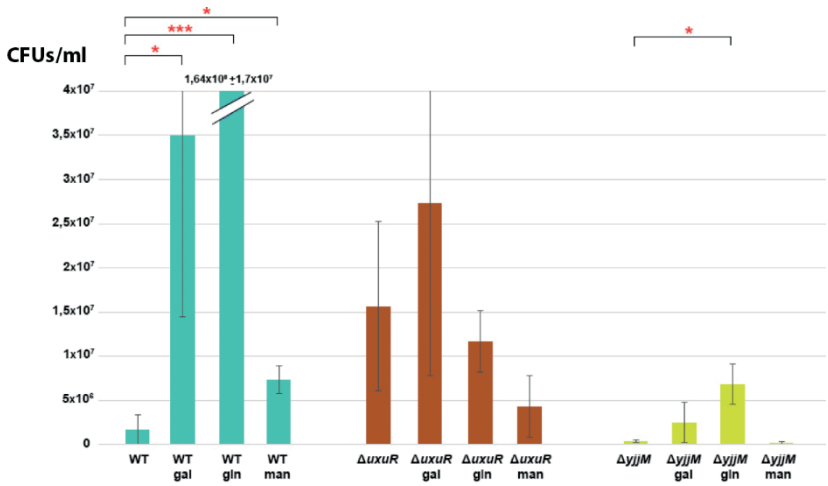


Fig. 1. Adhesion of *E. coli* K-12 and its deletion mutants for the *uxuR*, and *yjjM* genes in the presence of 0.2 % galacturonate (gal), glucuronate (gln), or mannose (man) in the medium to CaCo-2 cells. The standard deviation for three biological replicates is presented. Statistical significance was calculated using the t-test: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) relative to K-12 MG1655

YjjM indeed affected adhesion efficiency, while *uxuR* deletion led to the decrease of hexuronate-mediated adhesion activation (Fig. 1). Deletion of *crp* also decreased the number of attached cells, while deletion of *exuR* caused no significant effects (not shown). These results confirm the importance of hexuronate metabolism for successful surface attachment of bacteria to host tissues and the key role of UxuR and YjjM in this process. The next step was to evaluate the effect on the biofilm formation (Fig. 2). As expected, deletion of *yjjM* and *crp* resulted in decreased biofilm formation, either in the absence or presence of glucose, but the effect of the *uxuR* deletion was unstable. At the same time, deletion of *exuR* did not cause any effect, that was rather unexpected.
According to the transcriptomic data, the absence of uxuR resulted in dramatic activation of all fli – genes (responsible for flagella formation and motility), but only in two out of three cases. This explains the data on the proteomic maps of the K-12 and ΔuxuR in our earlier study (Fig. 4A) [2]. As such, the next part of our research was to verify or refute the assumption that the deletion of uxuR resulted in an imbalance in the transcription level of genes responsible for biofilm formation thus leading to different results on the plates that we had observed (Fig. 2). For this purpose, we studied biofilm formation and expression of genes encoding key regulators, in the individual colonies of E. coli K-12 MG1655, and its ΔuxuR, ΔexuR, Δcrp, and ΔyjjM derivatives. Since within uxuR small non-coding RNAs can be synthezised that are involved in the regulatory processes [3], two versions of the mutant were taken: full gene deletion and the strain where UxuR is not translated but all RNAs can be synthesized (ΔuxuR-tr). We measured expression of four genes: fliA coding for the “flagellar” sigma-factor σ28 [1]; csgD coding for the the pili responsible for the cell-surface adhesion [1]; csrC, small RNA that regulates csrA expression and was earlier shown to regulate biofilm formation [1]; and ssrS, 5S rRNA that was dramatically overexpressed in the biofilm-forming cells independent on the carbon source (Fig. 3).

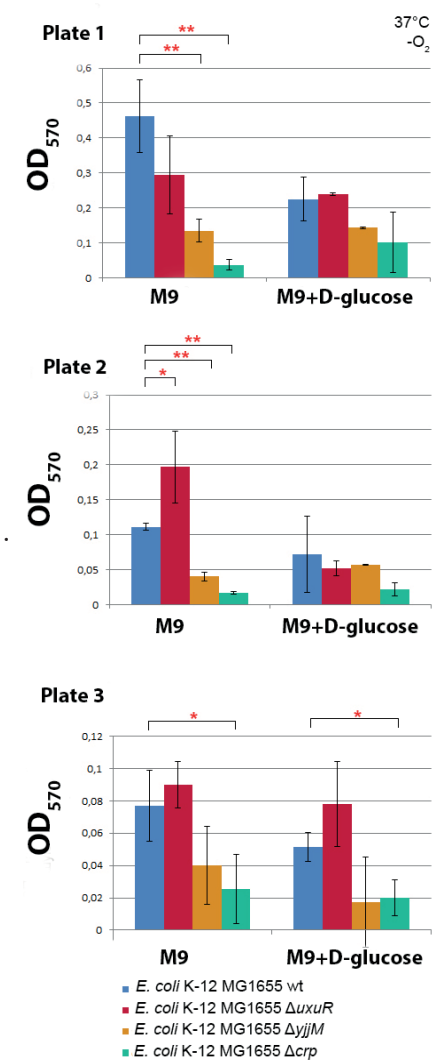


Fig. 2. Biofilm formation in 96-well plates. Data were normalized to the optical density. Data are presented for three identical plates grown simultaneously (biological repeats), and each sample was grown in three wells (three technical repeats). Cells were cultured for 48 hours in M9 + 5% LB + 0.2% D-glucose at 37 °C, microaerobic conditions. Spreads are presented as the standard deviation of three technical repeats. Statistical significance was calculated using t-test: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) relative to K-12 MG1655

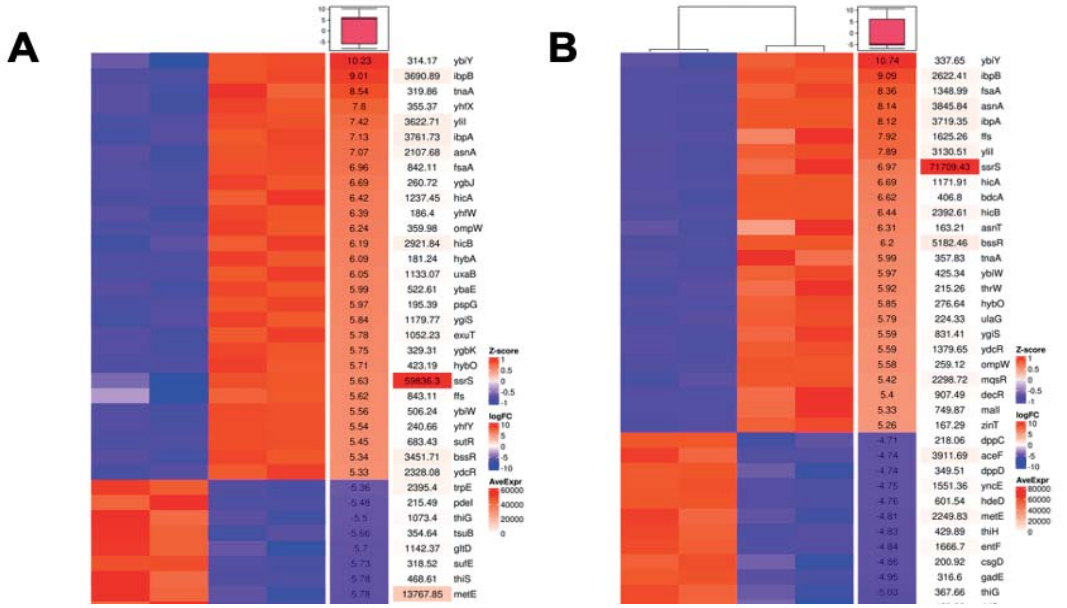


Fig. 3. Differential expression of genes for the free-living *E. coli* K-12 MG1655 and its biofilm during growth on D-glucose (A) and D-glucuronate (B)

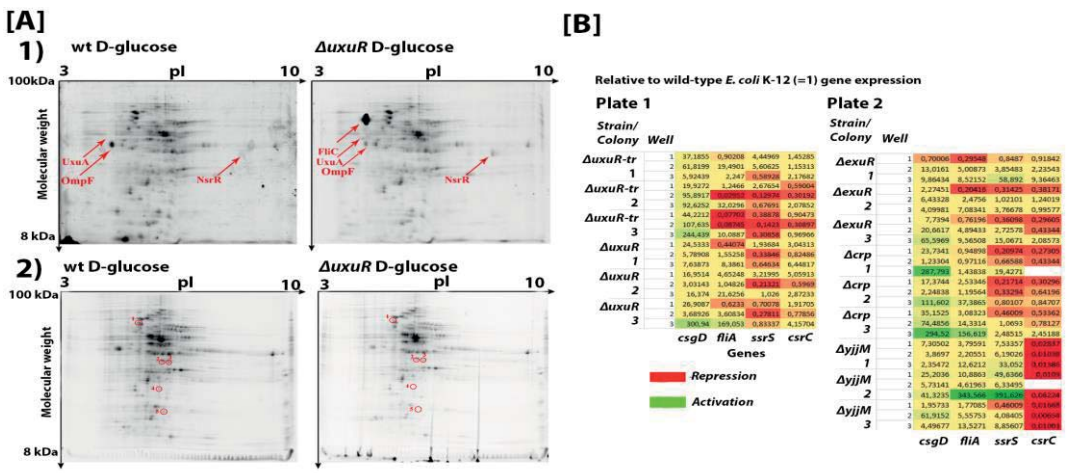


Fig. 4. [A] Proteomic maps of the wild type *E. coli* K-12 MG1655 and *E. coli* K-12 MG1655 *ΔuxuR* during growth with D-glucose: first experiment (1) and second experiment (2) [2]. [B] Gene expression of mutant strains presented as a heat map relative to *E. coli* K-12 (=1)

qRT-PCR (Fig. 4B) confirmed that transcription of fliA in ΔuxuR indeed appeared to be "floating", and a burst was observed only in several wells. In the ΔuxuR-tr mutant, fliA was repressed reflecting the impact of RNAs synthesized from within uxuR. The csgD expression was activated in both strains that corresponded to the efficiency of biofilm formation in the respective wells in this experiment. YjjM appeared to directly activate CsrC. Deletion of crp activated csgD and inhibited csrC, but the effects were less prominent than for deletion of yjjM. This is consistent with the previously shown CRP-mediated effect on csgD.

Conclusion: Hexuronates metabolism and its regulators, UxuR and YjjM, play the key roles in successful bacterial attachment and biofilm formation via modulating these processes, while cAMP-CRP probably act indirectly. Also, our data suggest that small non-coding RNAs are deeply involved in regulation of biofilm formation in *E. coli* K-12 and might serve as possible signaling molecules.

Funding: The study is supported by RSF 24-24-00435.

References

1. Salgado H., Gama-Castro S., Lara P. et al. RegulonDB v12.0: a comprehensive resource of transcriptional regulation in *E. coli* K-12. *Nucleic Acids Res*. 2024;52(D1):D255-D264. doi 10.1093/nar/gkad1072
2. Bessonova T.A., Fando M.S., Kostareva O.S., Tutukina M.N., Ozoline O.N., Gelfand M.S., Nikulin A.D., Tishchenko S.V. Differential impact of hexuronate regulators ExuR and UxuR on the *Escherichia coli* Proteome. *Int J Mol Sci.* 2022;23(15):8379. doi 10.3390/ijms23158379
3. Tutukina M.N., Dakhnovets A.I., Kaznadzey A.D., Gelfand M.S., Ozoline O.N. Sense and antisense RNA products of the *uxuR* gene can affect motility and chemotaxis acting independent of the UxuR protein. *Front Mol Biosci.* 2023;17(10):1121376. doi 10.3389/fmolb.2023.1121376

doi 10.18699/bgrs2024-1.2-10

Tissue-specific and allele-specific activity of transcribed gene regulatory elements in human skeletal muscle

Buyan A.I.[1, 3]*, Nozdrin V.A.[2], Gazizova G.R.[3], *Meshcheryakov* G.A.[1], Hayashizaki Y.[3], Shagimardanova E.I.[3, 4], Gusev O.A.[3, 4], Kulakovskiy I.V.[1, 3]

[1] *Institute of Protein Research, RAS, Pushchino, Russia*

[2] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

[3] *Research Laboratory "Regulatory Genomics", Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[4] *Life Improvement by Future Technologies (LIFT) Center, Moscow, Russia*

* *andreybuyanchik123@yandex.ru*

Key words: cap analysis of gene expression; transcribed regulatory elements; single-nucleotide variants; allele-specific activity; skeletal muscles

Motivation and Aim: The rapid development of sequencing technologies over the last decades brought molecular genetics close to its medical applications. However, genome-based personalized medicine is still in its early age as its implementation requires reliable tools for annotating and interpreting individual genome variants, including widespread single-nucleotide variants (SNVs). Hundreds of thousands of genome-wide association studies (GWAS) have been conducted to trace the links between genetic variants present in the human population and various phenotypes. Unfortunately, GWAS have limited resolution, strongly depend on the sample size to detect the associated variants, and do not directly differentiate between causal and linked variants. Post-GWAS functional annotation and identification of the causal variants remains challenging and critical for non-coding variants that usually comprise about 90 % of the GWAS credible sets [1]. Considering the regulatory non-coding regions, there are reliable maps of promoters and enhancers in immortalized cell lines and human tissues [2], but they do not describe the actual diversity of particular tissues, such as skeletal muscle. Among the current efforts in validation of regulatory variants detected in GWAS, skeletal muscle-associated phenotypes also remain among the least studied [1].

In this study, we utilized the cap analysis of gene expression (CAGE) data to construct an atlas of transcribed regulatory elements of diverse human skeletal muscles. Next, we assessed the functional effect of SNPs by estimating the allelic imbalance in the transcriptional activity of promoters and enhancers.

Methods and Algorithms: Here we used the results of 228 CAGE experiments from 76 muscles of 4 individuals. The basic bioinformatics pipeline included read mapping with HISAT2 [3], peak calling with CAGEfightR [4], SNV calling using bcftools and PySam [5], and estimating the allelic imbalance with MIXALIME [6].

Results: We identified more than 20 thousand CAGE clusters, including more than 1.5 thousand bidirectional enhancers. More than 3/4 of promoters and enhancers exhibited tissue-specific activity across different skeletal muscles, highlighting very specific gene expression patterns in eye, tongue, and diaphragm. The atlas yielded multiple interesting individual cases, such as tongue-specific keratin and bactericidal SPRR expression, or *CFAP61*, whose expression is positively correlated with aging [7] and was consistently low in eye samples. Furthermore, we compared the promoter

activity of skeletal muscles in different human body parts, considering Hox genes and myosin light and heavy chains, whose expression was in agreement with the muscle origin and functioning. Finally, we performed the differential promoter usage analysis to identify promoters driving different transcript isoforms depending on the muscle type. Among these genes, we found *SEPTIN9* that encodes a cytoskeleton protein containing microtubule-associated domain in most muscles, except for the eyes which use its shorter isoform [8]. Next, we performed the SNV calling which yielded more than 15 thousand reliable variants suitable for the allele-specific analysis. As a result, we detected more than 6 thousand allele-specific events (ASEs) with a significant allelic imbalance in at least one skeletal muscle type. More than half of these events were annotated as expression quantitative trait loci from GTEx [9] or allele-specific transcription factor binding sites from ADASTRA [10]. Besides the events that exhibited allele-specificity in some particular muscles, we also identified nearly thousand of SNVs demonstrating a significant difference in their allelic imbalance across the samples. Annotating these variants by phenotypes using stratified linkage disequilibrium score regression [11] revealed significant enrichment for the normalized muscle volume, suggesting the regulatory nature of the genetic basis of this trait [12].

Conclusion: Taken together, we constructed an atlas of transcribed regulatory elements in diverse human muscles (FANTOMUS: functional annotation of muscle promoters and enhancers), and performed the allele-specific analysis of promoters and enhancers. We believe that FANTOMUS will be useful for studying gene expression patterns and variant effects in healthy muscles as well as myopathies.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (No. 075-15-2021-601) and Russian Science Foundation (No. 20-74-10075).

References

1. Alsheikh A.J., Wollenhaupt S., King E.A. et al. The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. *BMC Med Genomics*. 2022;15(1):74. doi 10.1186/s12920-022-01216-w
2. Andersson R., Gebhard C., Miguel-Escalada I. et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507(7493):455-461. doi 10.1038/nature12787
3. Kim D., Paggi J.M., Park C. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907-915. doi 10.1038/s41587-019-0201-4
4. Thodberg M., Thieffry A., Vitting-Seerup K. et al. CAGEfightR: analysis of 5′-end data using R/Bioconductor. *BMC Bioinformatics.* 2019;20(1):487. doi 10.1186/s12859-019-3029-5
5. Danecek P., Bonfield J.K., Liddle J. et al. Twelve years of SAMtools and BCFtools. *GigaScience. 2021;10(2):giab008. doi 10.1093/gigascience/giab008*
6. *Meshcheryakov G., Abramov S., Boytsov* A. et al. MIXALIME: MIXture models for allelic imbalance estimation in high-throughput sequencing data. *arXiv*. 2024. doi 10.48550/arXiv.2306.08287
7. Tumasian R.A., Harish A., Kundu G. et al. Skeletal muscle transcriptome in healthy aging. *Nat Commun.* 2021;12(1):2014. doi 10.1038/s41467-021-22168-2
8. Kuzmić M., Linares G.C., Fialová J.L. et al. Septin-microtubule association via a motif unique to isoform 1 of septin 9 tunes stress fibers. *J Cell Sci.* 2022;135(1):jcs258850. doi 10.1242/jcs.258850
9. Lonsdale J., Thomas J., Salvatore M. et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585. doi 10.1038/ng.2653
10. Abramov S., Boytsov A., Bykova D. et al. Landscape of allele-specific transcription factor binding in the human genome. *Nat Commun.* 2021;12(1):2751. doi 10.1038/ng.2653
11. Finucane H., Bulik-Sullivan B., Gusev A. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet.* 2015;47(11):1228-1235. doi 10.1038/ng.3404
12. van der Meer D., Gurholt T.P., Sønderby I.E. et al. The link between liver fat and cardiometabolic diseases is highlighted by genome-wide association study of MRI-derived measures of body composition. *Commun Biol.* 2022;5(1):1271. doi 10.1038/s42003-022-04237-4

doi 10.18699/bgrs2024-1.2-11

Prediction and annotation of alternative transcription starts in the chicken genome

Grushina V.*, Pintus S.
Sirius University of Science and Technology, Sirius, Krasnodar region, Russia
** grushina_v@bk.ru*

Key words: CAGE; promotor shift; *Gallus gallus* genome; transcription start sites

Motivation and Aim: The analysis of promoter regions of genes often reveals differences from the existing annotation, which can be caused both by the insufficient accuracy of the existing annotation and the existence of transcripts previously unknown and activated under certain non-standard conditions [1, 2]. Taking into account the data from CAGE experiments would make it possible to better identify the initial positions of such transcripts, which allows us to determine the position of the start of transcription, as well as to identify the events of the promoter shift. Such insights would provide a better understanding of the mechanisms of transcriptional regulation in the chicken genome and improve the quality of Gallus gallus genome annotation.

Currently, data from CAGE experiments are available for various chicken tissues, including muscles, kidneys, liver, brain, and others. This provides the opportunity to annotate enhancers and alternative initial transcription positions that are active in various tissues. This approach ensures the completeness of the annotation, taking into account the differences in the gene expression in different tissues.

In this study, we predicted transcription start sites (TSS) in the chicken genome and evaluated their expression based on data obtained from CAGE experiments, including those of the FANTOM5 project [3]. Our predictions were compared against the Ensembl and Refseq annotations. We found new transcription starts, including those presumably associated with the events of promoter shift.

Methods and Algorithms: The galGal6 and galGal7 genome assemblies and Ensemble and Refseq genome annotations were used as a referencefor comparing and validating TSS coordinates. STAR [4] (v.2.7.11b) was used to align the CAGE reads against the reference genome, then the CAGE tag peaks were clustered with DPI1 tool [5]. The TSS coordinates and alternative transcription sites (promoter shift events) were identified with the CageFightR package [6] (v. 1.24.0).

Results: Differences were found between the predicted transcription starts in FANTOM5 and Ensembl and Refseq genome annotations of the chicken genome version 6, the re-aligned readings of FANTOM5 to the 7th version genome and Ensembl and Refseq genome annotations of the chicken genome version 7. Thus, the data from the FANTOM5 project were annotated to the latest version of the chicken genome, galGal7, through read realignment. This allowed the validation of TSS obtained as a result of the analysis of new CAGE experiments.

We predicted and annotated transcription sites in the chicken genome based on new CAGE experiments for different tissues. The CAGE reads were aligned to the galGal7 chicken genome and, based on the alignment results, the CAGE peaks were counted and clustered and noise corrected. The obtained transcription starts were compared by

interval arithmetic methods of the obtained transcription starts with Ensembl and Refseq annotations for the galGal7 chicken genome.
We identified and annotated statistically significant alternative transcription starts in the chicken genome (putative promoter shifts) based on data from the FANTOM5 project and new experimental CAGE data.
Conclusion: In addition to the major transcription starts annotated in Refseq and Ensembl reference annotations, CAGE experiments reveal additional TSS. Our results suggest that some genes are characterized by the phenomenon of transcription start shifts that manifest under specific conditions.
Funding: The study is by the Russian Science Foundation (project 24-24-20106, https://rscf.ru/project/24-24-20106/)

References

1. Link V.M. Analysis of genetically diverse macrophages reveals local and domain-wide mechanisms that control transcription factor binding and function. *Cell*. 2018;173:1796-1809. doi 10.1016/j.cell.2018.04.018
2. Anderson W.D., Duarte F.M. et al. Defining data-driven primary transcript annotations with primaryTranscriptAnnotation in R. *Bioinformatics*. 2020;36(9):2926-2928. doi 10.1093/bioinformatics/btaa011
3. Lizio M., Abugessaisa I., Noguchi S., Kondo A., Hasegawa A. et al. Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res*. 2019;47(1):752-758. doi 10.1093/nar/gky1099
4. Dobin A., Davis A.C., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi 10.1093/bioinformatics/bts635
5. Forrest A., Kawaji H., Rehli M. et al. A promoter level mammalian expression atlas. *Nature*. 2014;507(7493):462-470. doi 10.1038/nature13182
6. Thodberg M., Thieffry A., Vitting-Seerup K., Andersson R., Sandelin A. CAGEfightR: analysis of 5′-end data using R/Bioconductor. *BMC Bioinformatics*. 2019;20:487. doi 10.1186/s12859-019-3029-5

doi 10.18699/bgrs2024-1.2-12

Dormancy regulation of hematopoietic stem cell

Ibneeva L.[1], Singh S.P.[2], Sinha A.[1], Eski S.E.[2], Kovtun I.[1], Grinenko T.[1, 3]*

[1] *Institute for Clinical Chemistry and Laboratory Medicine, Technische Universität Dresden, Dresden, Germany*

[2] *IRIBHM, Université Libre de Bruxelles (ULB), Brussels, Belgium*

[3] *Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital Affiliated to Jiao Tong University School of Medicine, Shanghai, China*

** gri13101@rjh.com.cn*

Key words: hematopoietic stem cells; dormancy; CD38; c-Fos; $p57^{Kip}$

Motivation and Aim: Hematopoietic stem cells (HSCs) are responsible for the production of all blood cells during life. HSCs are very heterogeneous, while most adult HSCs are maintained in a quiescent state, numerous studies have demonstrated that 20–30 % of HSCs are deeply quiescent and, therefore, called 'dormant' HSCs (dHSCs). dHSC are characterized by slow metabolism, reduced ribosomal biogenesis, and DNA replication. Dormancy protects HSCs from external stresses, accumulation of somatic mutations, prevent their exhaustion and malignant transformation. Despite their deep quiescence, dHSCs harbor the greatest long-term repopulation capacity in transplantation assays. They do not produce cells under homeostatic conditions and are activated only in response to severe stress signals such as interferons, lipopolysaccharide, and myeloablation [1]. Thus, dHSCs serve as a reserve pool of stem cells throughout life. However, excessive dormancy may prevent of hematopoietic system from an efficient response to hematological stresses. Despite the importance of dormant HSCs, their detailed characterization has been challenging due to the absence of surface markers for their ready identification and isolation. Consequently, processes involved in preserving of dHSC quiescence are poorly understood. The aim of our research was to find a surface markers for isolation dHSC and to investigate molecular mechanisms of dormancy regulations.

Methods and Algorithms: we used single-cell and bulk next generation sequencing of murine HSCs. Pseudotime analysis, dormancy score calculation. Transplantation of HSCs into lethally irradiated recipients, FACS analysis and sorting, label retaining assay in vivo, single cell tracing assay in vitro, Ca2+ flux assay, and immunohistochemistry.

Results: To capture the transition between quiescence and proliferation, HSCs from young mice were subjected to single cell RNA sequencing, and after quality control, the transcriptome profiles of 1613 individual HSCs were used for downstream analysis. To identify actively cycling cells, we calculated cell cycle and dormancy scores of individual HSCs using Seurat [2], which were based on the expression levels of cell cycle and dormancy genes [3]. We observed that cells in the S and the G2/M phases were clustered together and that, as expected, most of the HSCs were quiescent (Fig. 1). Next, we attempted to isolate cell surface markers associated with HSC dormancy and identified *Cd38* as a putative marker for dHSCs because its expression was higher in cells with high dormancy scores and corresponded with the expression of well-known long-term HSC markers.

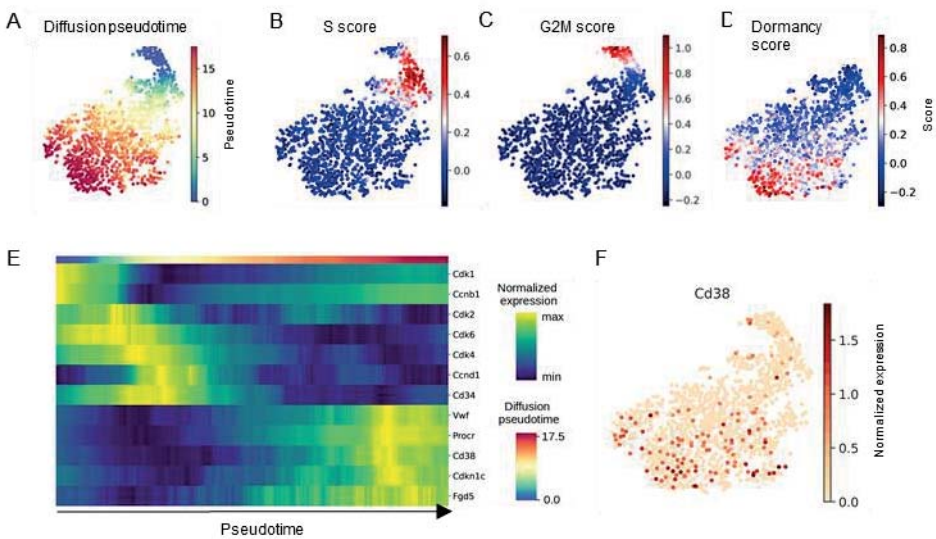


Fig. 1. Single cell transcriptome analysis of HSCs.
(A) Uniform manifold approximation projection (UMAP) representation depicting the transcriptional profiles of individual HSCs (LSK CD48^{-} CD150^{+}). (B) S-phase score along pseudotime. (C) G2/M phase score along pseudotime. (D) Dormancy score along pseudotime. For panels A–D, each dot represents a single cell. (E) Expression of selected genes along pseudotime. (F) UMAP representation showing the expression of *Cd38*. Each dot represents a single cell

We found that about 50 % of HSCs expressed CD38 on the surface at a high level. Using lable retaining assay together with cell cycle analysis of HSCs in vivo in steady state condition and in response to several hematopoietic stresses, we have confirmed that CD38+ HSCs represent the most quiescent population of stem cells. Moreover, CD38+ HSCs have higher repopulation capacity compared to their CD38- counterparts in serial transplantation assay. Therefore, our results indicate that CD38^{+} HSCs reside at the top of the hematopoietic hierarchy. Hence, CD38 represents a marker that will help circumvent the limitations of the long-term label-retaining assays or even negate the necessity of reporter mice for studying the mechanisms underlying HSC dormancy.
CD38 is a multifaceted nicotinamide adenine dinucleotide (NAD) catabolic ecto-enzyme that metabolizes NAD and its precursors (nicotinamide mononucleotide-NMN and nicotinamide riboside-NR) into adenosine diphosphate ribose (ADPR) and cyclic-ADPR (cADPR). We have found that inhibition of CD38 ecto-enzymatic activity led to the cell cycle activation of dHSCs. Moreover, we found that cADPR, the product of NAD conversion by CD38, activates Ca^{2+} release from endoplasmic reticulum in CD38+ dHSCs.
To clarify how the CD38/cADPR/Ca^{2+} axis regulates HSC dormancy, we performed a bulk transcriptome RNA sequencing of CD38+ and CD38- HSCs and found that while 205 genes were significantly down-regulated in CD38+ HSCs, 225 were up-regulated. Gene sets controlling the response to calcium ions, extracellular matrix interaction, and TGF-β1 response were up-regulated in CD38+ HSCs. HSC-related genes, such as *Hoxb9*, *H19*, *VwF*, *Clu*, and *Selp*, as well as genes associated with HSC dormancy, namely, *Gprc5c*, *Meis2*, and *Neo1*, were up-regulated in CD38+ HSCs. We did not find significant differences in *Cdk2*, *Cdk4*, *Cdk6*, and *CyclinD1* expression, but CD38+ HSCs

expressed the cell cycle inhibitors *Cdkn1a* and *Cdkn1c* at higher levels than CD38- cells. Intriguingly, the transcription factor *Fos*, whose expression was previously correlated to cell cycle activation, was one of the most significantly up-regulated genes in CD38+ HSCs. This observation was further corroborated by the fact that AP-1 complex responsive genes were enriched in CD38+ HSCs compared to CD38- counterparts. Moreover, $CD38^{+}$ HSCs displayed higher levels of transcriptionally active phosphorylated c-Fos (at Threonine 232, p-c-Fos) than CD38- cells. Using in vitro and in vivo assays, we found that inhibition of c-Fos transcription activity activated cell cycle entrance of dormant CD38+ but not CD38- HSCs. We identified that c-Fos regulates HSC dormancy via up-regulation of *Cdkn1c* expression ($p57^{kip2}$-cell cycle inhibitor). Importantly, we found that CD38 enzymatic activity on neighboring cells regulated the proliferation of CD38-negative human HSCs. In human bone marrow, several cell types express CD38: multipotent and restricted hematopoietic progenitors, plasma cells, activated T and B-lymphocytes, and NK cells. Therefore, some of these CD38+ cells can be the neighbors for human HSCs. Several hematological malignancies (chronic myeloid leukemia, acute myeloid leukemia, acute lymphoblastic leukemia, and multiple myeloma) express CD38 at a high level. We hypothesize that tumour microenvironment enriched in the products of CD38 ecto-enzymatic activity may keep healthy HSCs in the quiescent state leading to cancer-related pancytopenia as well as it may preserve the dormancy of cancer stem cells leading to disease persistence. Therefore, a better understanding of the mechanisms controlling human HSC dormancy is required to support healthy hematopoiesis in patients with hematologic malignancies and develop more powerful strategies for cancer eradication.

Conclusion: Here, we identify CD38 as a novel and broadly applicable surface marker for the enrichment of murine dHSCs. We reveal that the CD38/cADPR/Ca^{2+}/c-Fos/$p57^{kip2}$ axis regulates HSC dormancy. Mechanistically, we demonstrate that CD38 itself regulates stem cell dormancy by shuttling intracellular Ca^{2+} in a CD38/cADPR-dependent manner, which results in a consequent increase in c-Fos and the expression of the cell cycle inhibitor $p57^{kip2}$. Manipulation of this axis can potentially stimulate dHSC expansion and their efficient response to hematopoietic stress.

Funding: The study is supported Deutsche Forschungsgemeinschafft/DFG (No. GR. 4857/2-1).

References

1. Wilson A., Laurenti E., Oser G., van der Wath R.C., Blanco-Bose W., Jaworski M. et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell.* 2008;135(6):1118-1129. doi 10.1016/j.cell.2008.10.048
2. Stuart T., Butler A., Hoffman P., Hafemeister C., Papalexi E. et al. Comprehensive integration of single-cell data. *Cell.* 2019;177(7):1888-1902e21. doi 10.1016/j.cell.2019.05.031
3. Hao Y., Hao S., Andersen-Nissen E., Mauck W.M., Zheng S., Butler A. et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184(13):3573-3587e29. doi 10.1016/j.cell.2021.04.048

doi 10.18699/bgrs2024-1.2-13

Exploring the interplay between reproducibility of open chromatin regions and transcription factor functional activity

Kolmykov S.*, Kulyashov M., Sokolova T., Prasolov D., Kolpakov F.
Department of Computational Biology, Sirius University of Science and Technology, Sirius, Russia
** kolmykovsk@gmail.com*

Key words: transcription factors binding regions; open chromatin regions; ChIP-seq; DNase-seq; scATAC-seq

Motivation and Aim: Exploring the relationship between transcription factor binding regions (TFBRs) and open chromatin regions (OCRs) is crucial for comprehending gene expression regulation. Currently, there are thousands of publicly available next-generation sequencing (NGS) experiments on transcription regulation, such as ChIP-seq, DNase-seq, ATAC-seq, etc. Despite the enormous variety of experimental conditions, it is possible to identify groups of experiments conducted under the similar conditions. Hence, exploring the reproducibility of regulatory regions across experiments under the similar conditions alongside their functional analyses holds promise in advancing our understanding of transcriptional regulation mechanisms. The primary goal of this study is to assess the positional specificity of TFBRs relative to OCRs and their reproducibility, and to explore how this characteristic relates with the functional activity of TFBRs.

Methods and Algorithms: In this study, we utilized 1239 ChIP-seq and 500 DNase-seq experiments from the GTRD database [1]. Open chromatin regions (OCRs) and TF binding regions (TFBRs) were overlapped with each other within cell lines/tissues and under similar treatment conditions. OCRs were pooled and merged across cell lines, and for each region, its reproducibility within the cell line was evaluated. Sets of reliable TFBRs were selected using the METARA algorithm [2].

Functional annotation of TFs was carried out based on the Gene Ontology (GO) database. The TFs were divided into 3 groups: positive regulators (GO:0051254), negative regulators (GO:0051253), TFs with mixed activity. Additionally, TF knockdown/knockout RNA-seq datasets from the KnockTF database were utilized to assess the functional activity of TFBRs [3].

Position weight matrices (PWM) from the HOCOMOCO database [4] were used to identify TF binding motifs within TFBRs. The evolutionary conservativity of TFBRs was evaluated using the phastCons100way and phyloP100way tracks from the UCSC database.

To investigate the reproducibility of open chromatin regions at the single-cell level, we analyzed scATAC-seq data (GSE162690) using the ArchR R-library [5] and a homebrew R-script.

Results: To determine whether TFBRs tend to localize within OCRs, we overlapped TFBRs and OCRs from the GTRD database (Fig. 1). Notably, TFs such as ATF2, MAFK, MYB, SPI1, FOXA2, and CEBPB exhibited reduced ratios of TFBRs within OCRs. To gain deeper insight into this observation, we investigated whether these proteins have the capacity to influence chromatin accessibility. Pioneer and flanking accessibility-associated TFs were identified based on the study by Lemma et al. [6]. The

majority of TFs linked with chromatin remodeling exhibit a lower percentage of TFBRs within OCRs. Moreover, these TFs are characterized by lower reproducibility of OCRs.

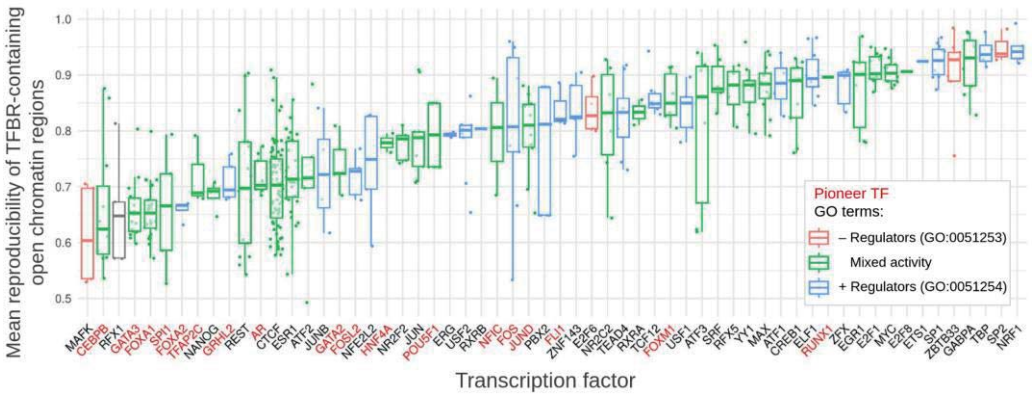


Fig. 1. The average reproducibility of OCRs overlapped with TFBRs in the cell type with the number of DNase-seq experiments greater than 4

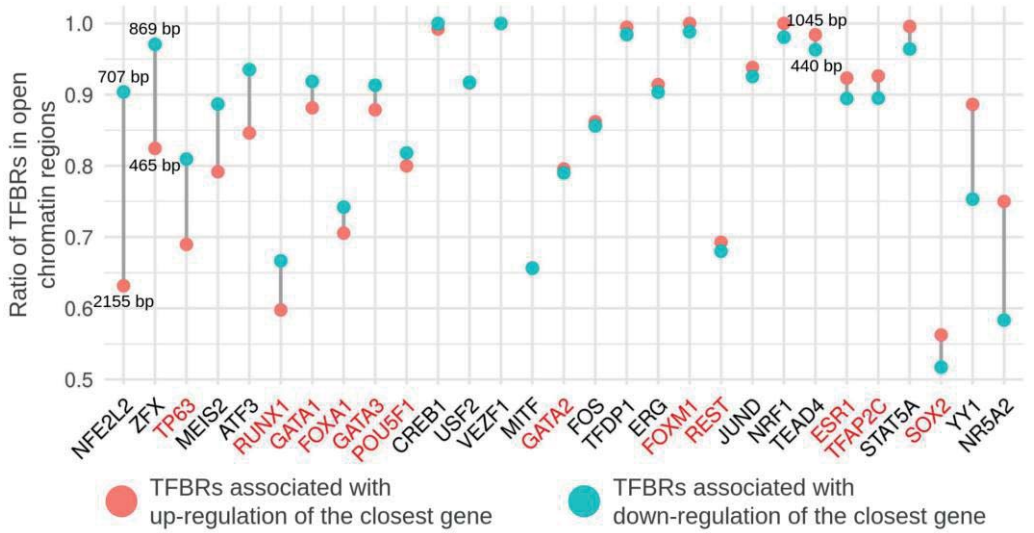


Fig. 2. Relationship between the ratio of TFBRs within OCRs and the pattern of change in the expression level of the closest gene following TF knockout/knockdown. Distances in kb indicate the average distance from the nearest TSS

Since DNase-seq is based on bulk cell analysis, it's likely that lower reproducibility of OCRs may arise due to the OCRs heterogeneity across different cell subpopulations. To explore this further we analyzed reproducibility of OCRs at the single cell level based on the scATAC-seq data.

Additionally, we investigated differences in the ratio of TFBRs within OCRs based on changes in the expression levels of the closest genes (<5kb to TSS) after TF knockout/knockdown. Significant differences in the ratio of TFBRs within OCRs were observed for several TFs depending on whether the expression of the closest gene increased or decreased (Fig. 2). For example, for ZFX and NFE2L2, a higher ratio of TFBRs within OCRs is associated with an increase in gene expression levels after TF knockout, while the reverse pattern is observed for TFAP2C and STAT5A.

Furthermore, a relationship between the distance of TFBRs to the closest TSS and the alterations in gene expression levels was discovered for several TFs: NFE2L2, ZFX and TEAD4. It is worth noting, that for NFE2L2 and ZFX there is also a significant difference between reproducibility of OCRs containing TFBRs associated with up-regulation and down-regulation of expression levels of genes. Moreover, for ZFX, the TFBRs associated with up-regulation of the closest genes are more likely to contain a TF binding motif and are also more evolutionary conserved compared to down-regulation associated TFBRs.

Conclusion: The preferences of TFBRs to be located within OCRs for different TFs were investigated. Most TFs associated with chromatin remodeling have demonstrated a reduced reproducibility of OCRs containing TFBRs. Certain TFs showed a correlation between a higher ratio of TFBRs within OCRs, their reproducibility, and changes in the expression levels of the nearest genes following TF knockout/knockdown. Analysis of OCRs reproducibility at the single-cell level revealed stable regions among OCRs with reduced reproducibility, consistent across cell subpopulations.

Funding: The study was financially supported by the Russian Science Foundation (project No. 24-24-20108, https://rscf.ru/en/project/24-24-20108/).

References

1. Kolmykov S. et al. Meta-analysis of ChIP-seq datasets through the rank aggregation approach. In: Cognitive Sciences, Genomics and Bioinformatics (CSGB-2020). Novosibirsk, 2020;180-184. doi 10.1109/CSGB51356.2020.9214614
2. Kolmykov S. et al. GTRD: an integrated view of transcription regulation. *Nucleic Acids Res*. 2021;D1:D104-D111. doi 10.1093/nar/gkaa1057
3. Feng C. et al. KnockTF 2.0: a comprehensive gene expression profile database with knockdown/knockout of transcription (co-) factors in multiple species. *Nucleic Acids Res*. 2024;52(D1):D183-D193. doi 10.1093/nar/gkad1016
4. Vorontsov I.E. et al. HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Res*. 2024;52(D1):D154-D163
5. Granja J.M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet*. 2021;53(3):403-411. doi 10.1038/s41588-021-00790-6
6. Lemma R.B. et al. Pioneer transcription factors are associated with the modulation of DNA methylation patterns across cancers. *Epigenet Chromatin*. 2022;15(1):13. doi 10.1186/s13072-022-00444-9

doi 10.18699/bgrs2024-1.2-14

From regulatory SNPs to pathways: multi-omics insights into the biological mechanisms underlying type 2 diabetes (TD2M) and metformin response

Korbolina E.E.*, Damarov I.S., Merkulova T.I.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** lungry@bionet.nsc.ru*

Key words: gene expression regulation; regulatory SNPs; GWAS phenotype association; type 2 diabetes mellitus; metformin treatment response

Motivation and Aim: The experience with GWAS analysis has revealed the success of studies in connecting genotype and phenotype but certain limitations when informing on the functional consequences of *SNPs* on genes. Therefore, the multi-omics studies aimed at identifying functionally significant genetic variants are gaining popularity. In this work, we focus on identifying regulatory SNPs (rSNPs) potentially involved in the development of type 2 diabetes mellitus (T2DM) and response to antihyperglycemic therapy.
Methods and Algorithms: Paired data on gene expression profiles (RNA-seq) and active chromatin marks H3K4me3 and H3K27ac (ChIP-seq) data were received for peripheral blood mononuclear cells (PBMCs) of nine healthy donors and analyzed to identify rSNPs associated with both allele-asymmetric binding and expression events [1]. GWAS catalog and Genotype-Tissue Expression (GTEx) project were used to explore the resulting rSNP list. We implemented RNA-seq datasets GSE221521 and GSE153315 to identify differentially expressed genes between T2DM and normal samples and between metformin responders and non-responders using R software. We then constructed protein-protein interaction networks using the STRING database, analyzed the networks by Cytoscape software, explored gene ontology and key pathway analysis and predicted *hub* genes targeted by identified rSNPs.
Results: We identified 14 796 rSNPs in the promoters of more than 5000 genes of human PBMCs and from this number 4280 rSNPs were associated with both GWAS-derived phenotypic traits and GTEx-reported eQTLs. 4612 DEGs were identified between T2DM patients with diabetic retinopathy and controls from GSE221521. The promoters of 1284 DEGs harbored rSNPs from our list. We found 31 hub genes in the PPI network in type 2 diabetes, including the genes involved in inflammation, obesity, insulin resistance, and signaling pathways related to the control of glucose levels (AMPK- and FoxO- signaling pathways by KEGG). In total, 496 DEGs were found between metformin responders and non-responders from GSE153315 and the promoters of 131 DEGs harbored identified rSNPs. Many of these were shown to be involved in the T2DM pathogenesis. The most significant group of DEGs targeted by rSNPs in metformin monotherapy response was a group of genes encoding known transcriptional regulators (including FOXP1, POU2F2, YY-1, KLF6, NFIX, and GMEB1 transcription factors).
Conclusion: We conclude that the list of human rSNPs identified from PBMCs add to the functional interpretation of GWAS-association signals. Our results also suggest the

candidate causal rSNPs for T2DM as well as for the beneficial effects of metformin and the rSNP-targeted genes are enriched in the pathways related to glucose metabolism and inflammation.

Funding: The study is supported by Russian Science Foundation (Grant No. 23-15-00113).

References

1. Korbolina E.E., Bryzgalov L.O., Ustrokhanova D.Z. et al. A Panel of rSNPs demonstrating allelic asymmetry in both ChIP-seq and RNA-seq data and the search for their phenotypic outcomes through analysis of DEGs. *Int J Mol Sci.* 2021;22(14):7240. doi 10.3390/ijms22147240

doi 10.18699/bgrs2024-1.2-15

Nuclear abundant stable intronic sequences RNA (sisRNAs) produced by a number of transcribed genes in chicken oocyte

Krasikova A.[1]*, Schelkunov M.[2, 3], Makarova N.[2], Fedotova A.[2, 4], Kulikova T.[1], Fedorov A.[1]

[1] *Saint-Petersburg State University, St. Petersburg, Russia*

[2] *Genomics Core Facility, Skolkovo Institute of Science and Technology, Moscow, Russia*

[3] *Institute for Information Transmission Problems, Moscow, Russia*

[4] *Lomonosov Moscow State University, Moscow, Russia*

* *alla.krasikova@gmail.com*

Key words: circular RNA; hypertranscription; lampbrush chromosomes; nuclear RNA-seq; oocyte nucleus; stable intronic lariats; stable intronic sequences RNA

Motivation and Aim: Birds, like many other vertebrates, have a hypertranscriptional type of oogenesis, with an exceptionally high rate of RNA synthesis in the oocyte nucleus. Higher RNA polymerase II velocity directly correlates with the production of newly synthesised circular RNA (circRNAs) [1]. Recently, we have undertaken a detailed characterisation of the nuclear and cytoplasmic transcriptome of chicken (*Gallus gallus domesticus*) oocytes at the lampbrush chromosome stage of oogenesis [2]. Here we aimed to characterise nuclear and cytoplasmic circRNAs including stable intronic sequences RNA.

Methods and Algorithms: Total RNA, including small RNA, was isolated from the nuclei and cytoplasm of growing chicken oocytes using TRIzol reagent, following a previously developed protocol [3]. To degrade linear RNA, total nuclear RNA was also treated with the 3' to 5' exoribonuclease R (RNase R). Next we performed strand-specific high-throughput sequencing of total nuclear and cytoplasmic RNA fractions, poly(A)-enriched RNA libraries, and small RNA fractions from lampbrush-stage oocytes. CIRCexplorer2 was used to identify circRNAs in total and poly(A)-enriched oocyte nuclear RNA samples and in total nuclear RNA after RNase R treatment. The programme "fast_circ.py" with the "de novo" option that comes with CIRCexplorer2 (Version 2.3.8) was used with default parameters for *de novo* prediction of circRNAs.

Results: In the oocyte nucleus, we detected transcripts for 10697 protein-coding genes and 2488 long non-coding RNA genes. In nuclear RNA fraction certain intronic sequences were predominantly abundant within transcribed genes most probably reflecting separated intronic RNAs. A higher coverage of the individual introns in total RNA-seq of nuclear RNA fraction appears as single or multiple peaks. Resistance to RNase R treatment indicates the lariat form of some of the intronic sequences. The RNAs of intronic origin thus represent stable intronic sequences RNA (sisRNAs), which have been previously characterised in *X. tropicalis* oocyte nuclei [4].

The appearance of mature mRNA for the genes that produce sisRNA suggests that they are spliced out of the nascent transcript. In fact, we checked that sisRNAs come from the same strand as pre-mRNAs. Furthermore, only certain intronic sequences are unusually stable, while the other introns of the same primary full-length transcript are degraded upon splicing. Using CIRCexplorer2, we predicted circRNAs *de novo* in RNA samples after RNase R treatment, in total and poly(A) RNA libraries, with low read

counts in poly(A) RNA libraries. Genomic regions covered by the identified circRNAs include 5'UTR, exonic and intronic regions. These data suggest that a number of circRNAs in the oocyte nucleus are also produced by the back-splicing mechanism and do not have an intron-lariat structure. 5'UTR-derived sisRNAs, either in linear or circular form, stored in the oocyte may be involved in regulating host gene expression. The nuclear abundant RNase R resistant sisRNA of the YWHAQ gene is an example of a sisRNA covering the 5'UTR. RNase R resistant, stable intronic lariats bearing snoRNA or scaRNA are also characterized. Such stable intronic lariats bearing snoRNA do not function as guide RNAs for rRNA and snRNA modification in contrast to mature snoRNA.

Conclusion: Comparison of the nuclear and cytoplasmic RNA profiles of chicken oocytes revealed the nuclear retention of certain introns in the form of sisRNAs resistant to RNase R treatment. The high number of certain intronic sequences stored within the oocyte nucleus is due to months of expression of corresponding genes on the lateral loops of lampbrush chromosomes, as well as due to their unusual stability. Notably, sisRNA production generally does not interfere with host gene expression and the appearance of the corresponding mRNA. We propose that nuclear retained sisRNAs, produced by a number of transcribed genes on the lateral loops of avian and amphibian lampbrush chromosomes via a splicing-dependent mechanism, can maintain the continued transcription of their host genes either at the lampbrush stage of oogenesis or during embryogenesis.

Funding: The research was supported by the Russian Science Foundation (grant #19-74-20075) and was performed using the equipment of the Genomics Core Facility (Skoltech) and Resource Center “Molecular and Cell Technologies” (Saint-Petersburg State University).

References

1. Zhang Y., Xue W. et al. The biogenesis of nascent circular RNAs. *Cell Rep*. 2016;15:611-624. doi 10.1016/j.celrep.2016.03.058
2. Krasikova A., Kulikova T., Schelkunov M., Makarova N., Fedotova A., Berngardt V., Maslova A., Fedorov A. The first chicken oocyte nucleus whole transcriptomic profile defines the spectrum of maternal mRNA and non-coding RNA genes transcribed by the lampbrush chromosomes. *bioRxiv*. 2024. doi 10.1101/2024.02.05.577752
3. Krasikova A.V., Fedorov A.V. Expression profiles of long and short RNAs in the cytoplasm and nuclei of growing chicken (*Gallus gallus domesticus*) oocytes. *Russ J Genet Appl Res*. 2016;6:307-313. doi 10.1134/S2079059716030059
4. Gardner E.J., Nizami Z.F., Talbot C.C., Gall J.G. Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. *Genes Dev*. 2012;26:2550-2559. doi 10.1101/gad.202184.112

doi 10.18699/bgrs2024-1.2-16

Bulged G-quadruplexes in the human genome: identification and characterization of a novel type of non-canonical G-quadruplex

Kuznetsov Vladimir A.[1], Papp Csaba[1], Mukundan Vineeth T.[2], Jenjaroenpun Piroon[3], Phan Anh Tuân[2, 4]

[1] *Department of Urology, Department of Biochemistry and Molecular Biology, SUNY Upstate Medical University, Syracuse, NY, USA*

[2] *School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore*

[3] *Division of Bioinformatics and Data Management for Research, Research Group and Research Network Division, Research Department, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand*

[4] *NTU Institute of Structural Biology, Nanyang Technological University, Singapore, Singapore*

Key words: G-quadruplex, Bulged G-quadruplex, Regulatory sites, human genome

Motivation and Aim: DNA sequence composition determines the topology and stability of G-quadruplexes (G4s). Bulged G-quadruplex structures (G4-Bs) are a subset of G4s characterized by 3D conformations with bulges. Current search algorithms fail to capture stable G4-B, making their genome-wide study infeasible. Here, we introduced three thermodynamically stable G4-B motif models representing a large family of computationally defined and experimentally verified potential G4-B forming sequences (pG4-BS).

Methods and Algorithms: We found 478 263 pG4-BS regions that do not overlap 'canonical' G4-forming sequences in the human genome and are preferentially localized in transcription regulatory regions including R-loops and open chromatin.

Results: Over 90 % of protein-coding genes contain pG4-BS in their promoter or gene body. We observed generally higher pG4-BS content in R-loops and their flanks, longer genes that are associated with brain tissue, immune and developmental processes. Also, the presence of pG4-BS on both template and non-template strands in promoters is associated with oncogenesis, cardiovascular disease and stemness. Our G4-BS models predicted G4-forming ability in vitro with 91.5 % accuracy. Analysis of G4-seq and CUT&Tag data strongly supports the existence of G4-BS conformations genome-wide. We reconstructed a novel G4-B 3D structure located in the E2F8 promoter.

Conclusion: pG4-B's G4-like conformation, strand-specificity, and physical association with gene transcription regulatory sites suggest that G4-B plays a role in many normal and pathogenic molecular processes. This study defines a large family of G4-like sequences, offering new insights into the essential biological functions and potential future therapeutic uses of G4-B.

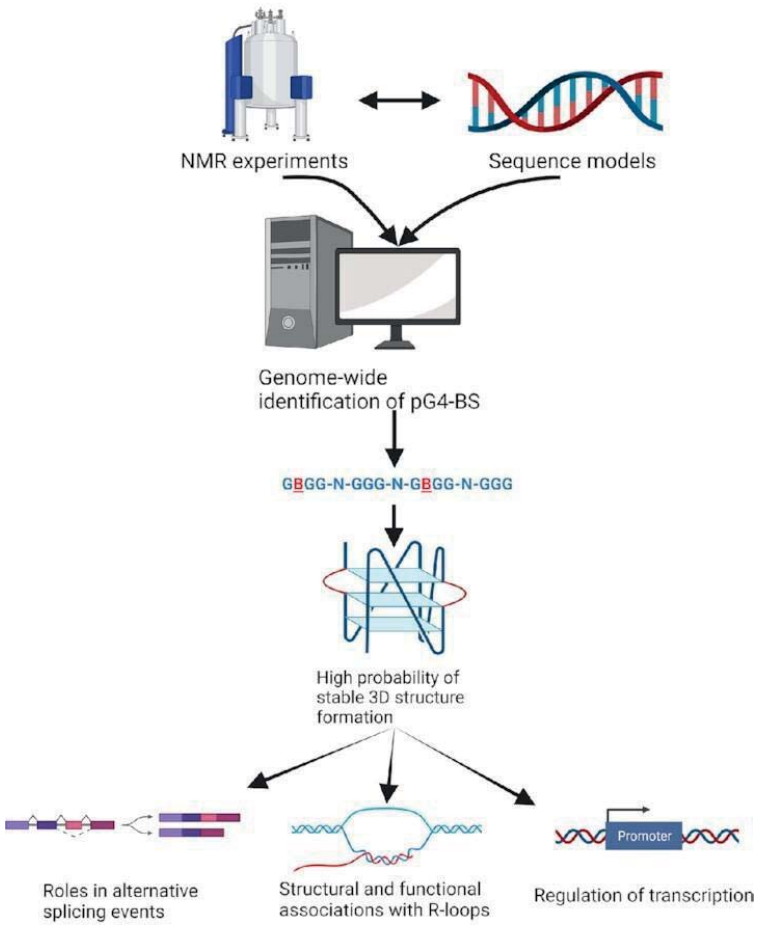


Fig. 1. This study computationally identified and experimentally validated a large family of G-quadruplexes-like sequences, called stable bulged G4, offering new insights into the genome and gene functions and future therapeutic uses

Funding: V.A.K. was supported by a SUNY EMPIRE innovation program scholar grant; Upstate Medical University Cancer Center grant and Upstate Foundation Turn4ACure Fund; Nanyang Technological University Singapore supported research in A.T.P. lab; P.J. was supported by the Office of the Permanent Secretary, Ministry of Higher Education, Science, Research and Innovation (OPS MHESI); Thailand Science Research and Innovation (TSRI) [RGNS 64-161].

References

1. Papp C., Mukundan V.T., Jenjaroenpun P., Phan A.T., Kuznetsov B.A. Stable bulged G-quadruplexes in the human genome: identification, experimental validation functionalization. *Nucleic Acids Res*. 2023;51(9):4148-4177. doi 10.1093/nar/gkad252
2. Papp C., Jenjaroenpun P., Mukundan V.T., Phan A.T., Kuznetsov V.A. Dataset of bulged G-quadruplex forming sequences in the human genome. *Data Brief*. 2023;50:109550. doi 10.1016/j.dib.2023.109550

doi 10.18699/bgrs2024-1.2-17

MetArea tool for predicting structural variability and cooperative binding of transcription factors in ChIP-seq data

Levitsky V.G.[1, 2]*, Tsukanov A.V.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *levitsky@bionet.nsc.ru*

Key words: transcription factor binding site motifs; performance of motif recognition; composite elements; recognition performance of motif models

Motivation and Aim: The rapidly growing amount of data on chromatin immunoprecipitation followed by massive sequencing (ChIP-seq) facilitates the search for the motifs of transcription factors (TFs) specific to a given cell or tissue type and condition [1]. Typically, two or more TFs work together to induce transcriptional changes [2]. However, progress in the development of methods for identifying the motifs of TFs involved in functionally important co-operative interactions still lags far behind the development of tools for finding the motifs of individual TFs.

Methods and Algorithms: For the model of TF binding site motif recognition, we calculate the recognition accuracy as the partial area under ROC (Receiver Operating Characteristic) curve (pAUC, [3]). Generalizing the ROC curves and pAUC values of the two separate models to a single joint model reveals whether the joint performance of two motifs competes with those of the participating motifs. Testing multiple possible combinations of motifs reveals the motifs most strongly reinforcing each other. These pairs of motifs can represent either structurally different types of binding sites for the same TF, or binding sites of different TFs acting as part of a single multi-protein complex. Thus, MetArea predicts motifs with synergistically related functions in gene transcription regulation. The synergy criterion requires for the pair of motifs 1 and 2 the higher pAUC1&2 value of the joint model 1&2 compared to the pAUC values of both participating motif models pAUC1 and pAUC2, Ratio = pAUC(1&2) / Max(pAUC1, pAUC2) > 1. MetArea considers ready ROC curves for TF binding site motif models, thereby it is applied to the motifs of the traditional model (Position Weight Matrix, PWM) or an alternative one, e.g. SiteGA [3].

First, MetArea algorithm computes the ROC curves for both motif models. For each motif model this curve represents the relationship between the rates of true positive (TP) and false positive (FP) predictions. For each recognition threshold, the TP rate denotes the fraction of nucleotide sequences from the foreground set (ChIP-seq peaks) containing predicted sites of a motif model. For each recognition threshold, the FP rate is defined as the frequency of the motif in the background set of nucleotide sequences [3]. The background set means genomic sequences with A/T content matching that for the sequences from the foreground set. Preliminary, for each motif model, the promoter sequences of protein coding genes from whole genomes are used to compute – Log_{10}(ERR) values (Expected Recognition Rate, ERR). These ERR values measure various motif models in the uniform scale [3, 4]. Thus, in terms of the FP rate computed through ERR values, for any pair of sites of two models, we can estimate which of two sites respects higher affinity.

Next, MetArea algorithm computes the ROC curves for the combination of two motif models. For them, the joint TP rate is the fraction of sequences containing predicted sites of at least one of two motif models. The joint FP rate requires the correction since predicted sites of two motif models may overlap each other in a nucleotide sequence. Only the three distinct cases are possible in one DNA sequence for two sites of two motif models: two sites do not overlap; they have full overlap, i.e. one of them is located entirely inside the other; they have partial overlap, i.e. neither one can be completely located within another. Correspondingly, MetArea corrects the FP rate for the joint motif model as follows. If two sites do not overlap, then the joint FP rate increases by two, $\Delta FP = 2$. If two sites have full overlap, then the joint FP rate increases by one, $\Delta FP = 1$. If two sites have partial overlap, then the joint FP rate increases by the non-integer impact between one and two ($1 < \Delta FP < 2$), it depends on the ratios between the overlap length (L) of two motifs, and the lengths of motifs W_1 and W_2,

$$\Delta FP = \frac{1}{2}\left(\frac{L}{W_1} + \frac{L}{W_1}\right) = \frac{L}{2}\left(\frac{W_1 + W_1}{W_1 * W_1}\right)$$

The cases of full or partial overlaps imply that the joint FP rate possesses the best ERR score among the two sites,

$$-Log_{10}(ERR_{1\&2}) = Max\{-Log_{10}(ERR_1), -Log_{10}(ERR_2)\}$$

Results: We took in analysis the ChIP-seq dataset of 1000 top-scored peaks for BHLHA15 TF from GTRD [5], ID PEAKS039234, GEO ID GSE86289 [6], pancreas of adult mice. We performed *de novo* motif search with the traditional PWM [7], categorized [1, 8] seven top-scored motifs to the motifs of known transcription factors and sorted these motifs by the pAUC measure (Table 1). The first two and third motifs represent BHLA15 binding sites, E-boxes of GC and AT types [6], correspondingly; the rest motifs are designated by TF names. We approved that the joint models {BHLHA15_GC_1 & BHLHA15_AT}, {BHLHA15_GC_2 & BHLHA15_AT} and {PRD13 & FOXA2} achieved an increase in accuracy of about 20 % compared to the respective single models (Table 1, Fig. 1A). Thus, the first two pairs of motifs represent the combinations of GC and AT E-boxes motifs for BHLA15 (Table 1). For the first pair pAUC values for single motifs are 7.45E-4 and 4.48E-4, while the joint model shows the pAUC value of 8.91E-4 (Fig. 1B).

Table 1. Seven top-scored motifs for PEAKS039234 dataset for murine BHLHA15 TF found by *de novo* motif search [7]. For each motif its name [5, 6], logo [8] and the performance pAUC are shown

Motif	Logo	pAUC, ×10000
BHLHA15_GC_1		7.45
BHLHA15_GC_2		6.45
BHLHA15_AT		4.48
CTCF		3.31
PRD13		3.23
FOXA2		2.66
GATA4		0.73

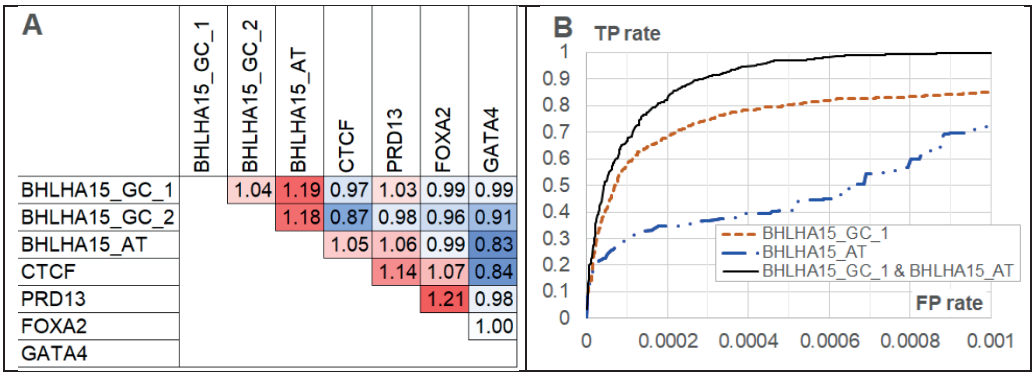


Fig. 1. Application of MetArea tool for the dataset PEAKS039234 for murine BHLHA15 TF from GTRD [5], GEO ID GSE86289 [6]. **A.** Ratio values for all 21 possible pairwise combinations of seven motifs from *de novo* motif search (Table 1, [7]). Shades of red/blue mark the Ratio values with/without the synergy of motifs. **B.** ROC curves for the motifs BHLHA15_GC_1, BHLHA15_AT (non-solid red and blue lines), and their joint model (solid black) achieving the higher performance

Conclusion: MetArea propose an approach to clarify the participants of multiprotein complexes regulating gene transcription. It can be used either to confirm distinct structural types of binding sites for the same TF, or to reveal the motif co-occurrence respecting to composite elements formed by the same or distinct motifs, homo- and heterotypic composite elements, correspondingly. Besides, MetArea tool can point to the false positive results of *de novo* motif search, since for the functional motifs the synergetic growth of the performance pAUC in the pairwise joint models is expected.
Funding: The study is supported Russian Science Foundation project No. 20-14-00140.

References

1. Vorontsov I.E., Eliseeva I.A., Zinkevich A., Nikonov M., Abramov S., Boytsov A. et al. HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Res.* 2024;52(D1):D154-D163. doi 10.1093/nar/gkad1077
2. Morgunova E., Taipale J. Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol.* 2017;47:1-8. doi 10.1016/j.sbi.2017.03.006
3. Tsukanov A.V., Mironova V.V., Levitsky V.G. Motif models proposing independent and interdependent impacts of nucleotides are related to high and low affinity transcription factor binding sites in Arabidopsis. *Front Plant Sci*. 2022;13:938545. doi 10.3389/fpls.2022.938545
4. Levitsky V., Zemlyanskaya E., Oshchepkov D., Podkolodnaya O., Ignatieva E., Grosse I. et al. A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package. *Nucleic Acids Res*. 2019;47(21):e139. doi 10.1093/nar/gkz800
5. Kolmykov S., Yevshin I., Kulyashov M., Sharipov R., Kondrakhin Y., Makeev V.J. et al. GTRD: an integrated view of transcription regulation. *Nucleic Acids Res.* 2021;49(D1):D104-D111. doi 10.1093/nar/gkaa1057
6. Jiang M., Azevedo-Pouly A.C., Deering T.G., Hoang C.Q., DiRenzo D., Hess D.A. et al. MIST1 and PTF1 collaborate in feed-forward regulatory loops that maintain the pancreatic acinar phenotype in adult mice. *Mol Cell Biol*. 2016;36(23):2945-2955. doi 10.1128/MCB.00370-16
7. Bailey T.L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics*. 2021;37:2834-2840. doi 10.1093/bioinformatics/btab203
8. Gupta S., Stamatoyannopoulos J.A., Bailey T.L., Noble W.S. Quantifying similarity between motifs. *Genome Biol*. 2007;8(2):R24. doi 10.1186/gb-2007-8-2-r24

doi 10.18699/bgrs2024-1.2-18

Prediction of expression changes in single cell RNA using style transfer variational autoencoder

Markelova E.[1, 2]*, Antonets D.[1], Minin A.[1], Vyatkin Y.[1], Shtokalo D.[1], Medvedeva Y.[3], Golovkin A.[4], Kostareva A.[4], Ramensky V.[1, 5, 6]

[1] *Institute for Advanced Research on Artificial Intelligence and Intelligent Systems, Lomonosov Moscow State University, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, Moscow, Russia*

[3] *Federal Research Centre "Fundamentals of Biotechnology", RAS, Moscow, Russia*

[4] *Almazov National Medical Research Centre, St. Petersburg, Russia*

[5] *National Medical Research Center for Therapy and Preventive Medicine of the Ministry of Healthcare of Russian Federation, Moscow, Russia*

[6] *Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

* *markelova.ke@gmail.com*

Key words: single cell sequencing; perturbation; deep learning; variational autoencoder; style transfer

Motivation and Aim: Single cell transcriptomics is already a well-established method for unbiased profiling of complex and heterogeneous systems [1]. The produced data is usually used to explain phenotypes based on the composition and dynamics of cells. The dynamics of individual cells in response to perturbations, like treatments or gene knockouts, are of particular interest. While the discovery of genes affected by a perturbation has been made possible by developments in single cell differential expression analysis [2], generative modeling of perturbation response goes one step further since it allows generating data *in silico*. The ability to generate data that was previously unseen (not present in the training data, out-of-sample) is particularly challenging and enticing.

Methods and Algorithms: In this study, we've polished the architecture and evaluated the performance of stVAE [3], a deep learning model that uses Conditional Variational Autoencoder [4] and Y-Autoencoder [5] training approaches. The stVAE architecture allows to disentangle the variables representing a needed 'style' from the hidden variables. We've applied stVAE to the single cell dataset of peripheral blood mononuclear cells (PBMCs) derived from patients with pericarditis, some of which received IL-1β blocker treatment. We've analyzed the distribution and composition characteristics of the predicted data, and also performed differential expression and pathway analyzes.

Results: Our goal was to train the model to predict treatment effects for specific cell types. As a first test, we trained the model on all the cell types present in the dataset and analyzed the predictions for the validation set of cells, left out during training. Then we also experimented by restricting the training data to a subset of cell types (e.g. training only on B cells, CD4 T cells, etc.) and then validating on the cell types hold out during training. Results from both approaches showed high correlations of predicted gene expressions and ground truth (mean R2 over all genes in cell types ≥0.93, mean R2 over differentially expressed genes in monocytes ≥0.87). We've proposed several metrics to evaluate the accuracy of perturbation predictions in comparison to the ground truth

(experimental data), for example, correlation of differentially expressed gene sets or fold changes and intersection or semantic similarity of enriched pathways.
Conclusion: We have demonstrated that stVAE is able to accurately predict single-cell perturbation response out-of-sample. Obtained predictions are cell-type- and patient-specific. To enable this, the model is required to capture features that differentiate genes and cells that respond weakly from those that respond strongly. Building biological interpretations of the features could, for instance, aid in comprehending how patients react to medications and thus advancing the personalized treatments.

References

1. Angerer P. et al. Single cells make big data: New challenges and opportunities in transcriptomics. *Curr Opin Syst Biol*. 2017;4:85-91
2. Nguyen H.C.T. et al. Benchmarking integration of single-cell differential expression: 1. *Nat Commun.* 2023;14(1):1570. doi 10.1038/s41467-023-37126-3
3. Russkikh N. et al. Style transfer with variational autoencoders is a promising approach to RNA-Seq data harmonization and analysis. *Bioinformatics*. 2020;36(20):5076-5085
4. Sohn K., Lee H., Yan X. Learning structured output representation using deep conditional generative models. In: Advances in Neural Information Processing Systems (NIPS 2015). 2015;1-9
5. Patacchiola M., Fox-Roberts P., Rosten E. Y-Autoencoders: Disentangling latent representations via sequential encoding. *Pattern Recognit Lett*. 2020;140:59-65

doi 10.18699/bgrs2024-1.2-19

Small non-coding RNAs of extracellular vesicles of parasitic worms as regulators of human target genes

Pakharukova M.Y.[1, 2]*, Lishai E.A.[1, 2], Medvedeva E.V.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *pmaria@yandex.ru*

Key words: microRNA; exosomes; transcriptome; human cells

Motivation and Aim: Small non-coding RNAs, including microRNAs, regulate gene expression through interaction with the 3'-UTR of target mRNAs, affecting their stability and/or translation. Parasitic organisms also secrete microRNAs in extracellular vesicles (EVs), which are able to penetrate human cells and modulate cellular processes [1–3]. The identification of such therapeutic parasitic microRNAs is especially important because many parasitic infections are negatively correlated with the morbidity and symptoms of autoimmune inflammatory diseases such as Asthma or Crohn's disease.

The trematode *Opisthorchis felineus* parasitizes the bile ducts of fish-eating mammals, including humans. Human infection leads to biliary epithelial neoplasia, on the other hand, to an alleviation of Asthma symptoms. The mechanisms by which parasites alleviate acute inflammation, regulate the immune response, and stimulate the neoplasia development have not been elucidated, in particular, the role of parasitic extracellular vesicles and non-coding RNAs has not been studied.

Aim: Identification of small non-coding RNAs of EVs of *Opisthorchis felineus* trematodes that exhibit specific bioregulatory activities in relation to human cells.

Methods and Algorithms: UMI-containing DNA libraries from small RNA of *O. felineus* EVs were constructed using the MGIEasy Small RNA Library Prep Kit V2.0 and sequenced 1X50 bp on BGISEQ-500 (BGI, China). DNA-libraries to study mRNAseq of human H69 cholangiocytes were constructed using the NEBNext Ultra II Directional RNA Library Prep Kit, the libraries were sequenced 2X150 base pairs on DNBseq (BGI, China). The quality of the data after sequencing was controlled using the FastQC program. Mapping of small RNA libraries to the reference genome of *O. felineus* was carried out using bowtie2. The miRDeep2 algorithm was used to identify known microRNAs. The list of known trematode microRNAs and their precursors was compiled according to the paper [4]. The conservation of major EVs microRNAs was studied using miRBase database. To search for target genes, a combination of TargetScan and miRDB was used.

To analyze the mRNAseq data from human cholangiocytes (H69) after exposure to extracellular vesicles of *Opisthorchis felineus*, sequence mapping was carried out on the human genome GRCh38 using STAR algorithm (v.2.7.10b). To analyze differential gene expression, the Wald test with a threshold of 0.05 from the DESeq2 (v.1.42.0; R package) was used. A correction for multiple comparisons (Benjamini-Hochberg) was applied to the p-values obtained, and genes were identified as differentially expressed (DEGs) when Padj <0.05 and a change in gene expression of more than 1.5 times. Principal component analysis to assess the degree of library clustering was carried out using PCAtools (v.2.14.0, R-package). Pathway enrichment analysis was performed

using Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Molecular Signatures Database (MsigDB) databases and msigdb (v.1.10.0), as well as clusterProfiler (v .4.10.0) R-packages.

Results: After sequencing of DNA libraries of small RNAs from EVs, an average of 3 million reads per library was obtained. 97 mature microRNAs were found in EVs. The most represented microRNAs were Ofe-Mir-277-P2_3p (18.5 %), Ofe-Mir-71-P1_5p (17.7 %), Ofe-Bantam_3p (11.8 %), Ofe-Mir-10 -P1_5p (11.3 %), Ofe-Mir-219_5p (6.0 %), Ofe-Mir-7-P1_5p (3.5 %). Differences were identified in the relative abundance of miRNAs in EVs compared to that of adult whole worms. Thus, the most represented microRNA in adult whole worms was Ofe-Mir-10-P2a (48 %), while in EVs the content of this microRNA was 4.41 %. In contrast, the most abundant miRNA in exosomes, Ofe-Mir-277-P2_3p, was presented less than 0.02 % in adult worms. Thus, we observed an enrichment of exosomes with certain microRNAs, which may indicate their functionality during interaction with host cells.

Among the major miRNAs of *O. felineus* EVs, there were both conserved miRNAs and those specific to parasitic worms. For example, seed-region of Ofe-Bantam-3p was identical only in those from Platyhelminthes and Nematoda species. In contrast, Ofe-Mir-10-P1 and Ofe-Mir-10-P2a are highly conserved and are also present in humans (Mir-125) with an identical seed-region. It can be assumed that *O. felineus* microRNAs can act on the same target genes as their human homologues.

Among the predicted target genes in the human genome for Ofe-Mir-277-P2, enrichment was observed in the genes of epithelial-mesenchymal transition pathway (p-value = 0.00063), protein secretion (p-value = 0.016), myogenesis (p-value = 0.046), PI3K-Akt signaling pathway (p-value = 0.041) and regulation of the actin cytoskeleton (p-value = 0.041). For Ofe-Mir-10-P1, enrichment among human target genes was observed in the pathway associated with proteoglycans in carcinogenesis (p-value = 0.00374).

As a result of sequencing the transcriptome of human H69 cholangiocytes (6 libraries), an average of 71 million reads per library were obtained. Analysis of transcriptomes revealed 773 differentially expressed genes (DEGs) after treatment of cells with trematode EVs (expression changes more than 1.5 times, Padj < 0.05), with 365 down-regulated genes and 408 upregulated genes. The Fisher test showed the enrichment of cholangiocyte DEGs with predicted target genes of parasitic microRNAs.

Conclusion: Thus, we have demonstrated the data indicating the regulatory role of parasitic microRNAs in the regulation of human gene expression. The study of parasitic secretory microRNAs provides the opportunity to identify new anti-inflammatory and biologically-active compounds created during co-evolution of the parasite and the host to modulate the immune response or stimulate the regeneration in tissues.

Funding: The study is supported by the Russian Science Foundation (No. 24-44-00048).

References

1. Pakharukova M.Y., Savina E., Ponomarev D.V., Gubanova N.V., Zaparina O. et al. Proteomic characterization of Opisthorchis felineus exosome-like vesicles and their uptake by human cholangiocytes. *J Proteomics*. 2023;283-284:104927. doi 10.1016/j.jprot.2023.104927
2. Ponomarev D.V., Lishai E.A., Kovner A.V., Kharkova M.V., Zaparina O. et al. Extracellular vesicles of the liver fluke Opisthorchis felineus stimulate the angiogenesis of human umbilical vein endothelial cells. *Curr Res Parasitol Vector-Borne Dis*. 2023;7(4):100153. doi 10.1016/j.crpvbd.2023.100153
3. Meningher T., Barsheshet Y., Ofir-Birin Y., Gold D., Brant B. et al. Schistosomal extracellular vesicle-enclosed miRNAs modulate host T helper cell differentiation. *EMBO Rep.* 2020;21(1):e47882. doi 10.15252/embr.201947882
4. Ovchinnikov V.Y., Mordvinov V.A., Fromm B. extreme conservation of miRNA complements in opisthorchiids. *Parasitol Int*. 2017;66:773-776

doi 10.18699/bgrs2024-1.2-20

Transcription factors with different DNA-binding domains respect different hierarchic levels such as class, family and subfamily as sufficient for significant similarity of their motifs

Raditsa V.V.[1]*, Tsukanov A.V.[1], Levitsky V.G.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *radicav06@gmail.com*

Key words: transcription factor binding sites; analysis of ChIP-seq data; de novo motif search; similarity of motifs; classification of transcription factors by their DNA-binding domains

Motivation and Aim: Transcription factors (TFs) are proteins that regulate transcription by binding to genomic DNA. TF binding sites (TFBS) in genomic DNA usually varied in length from 6 to 20 bp, their common pattern for particular TF is called motif. The ChIP-seq experimental technology, based on chromatin immunoprecipitation (ChIP), allows genome-wide mapping of genomic loci (peaks) potentially containing these sites for a target TF. The important task of ChIP-Seq data analysis is to detect enriched motifs in peaks. A *de novo* motif search applied to sets of foreground (peaks) and background (non-specific DNA) sequences can learn enriched motifs that represent an exact model of a target TF motif recognition [1]. The significance of similarity of the motifs of known TFs [2] from public databases [3, 4] to enriched motifs supports their relation to particular TFs. For each TF, the structure of its DNA-binding domain (DBD) defines available variability of its motif. The hierarchical classification of TFs by DBD [5] defines top two levels as superclasses and classes based on the general type of DBD and its structure. Next two levels, families and subfamilies are defined by sequence similarity of DBDs. The main idea here is to trace gradually the connection between similar DBDs and related DNA sequences. However, this classification does not have strict criteria for the significance of similarity of TFBS motifs at certain levels of TF classification. Identification of TFBS motifs is complicated by the noticeable increase in the number of TFs with known motifs (there are already more than a thousand of them [3, 4]). However, the number significantly different TFBS motifs is substantially less, since the motifs of structurally related TFs may be practically indistinguishable [6]. For some classes, there may be two or more structurally different DNA motifs per TF. E.g., TFs can bind only as dimers, or as monomers or dimers, etc. [7].

To date, for TFs of a given DBD structure (class, family, and subfamily) it has not been determined which of these levels is sufficient to identify a set of TFs with significantly similar motifs. The aim of this study is to include the systematic analysis of similarity of motifs of known TFs from different levels of TF classification to the standard protocol of *de novo* motif search applied to the results of ChIP-seq experiment. We performed the massive analysis of ChIP-seq data, estimated the similarity of motifs of different target TFs within certain classes/families and compared these distributions with those computed for distinct families within each class.

Methods and Algorithms: We extracted ChIP-seq datasets for *M. musculus* from the GTRD [8]. We used the hierarchical classification of TFs by DBDs into classes,

families [5]. Six most numerous TF classes (Table 1) were included in the analyses. We used *de novo* motif search tool STREME [1], it applies the traditional motif model of Position Weight Matrix. We applied TOMTOM tool [2] (a) to confirm the significant similarity of enriched motifs to motifs of known TFs from JASPAR/Hocomoco [3, 4] databases, and (b) to estimate the similarity in pairs of enriched motifs. To define the distributions of significance motifs similarity within a class or family, we used only pairs of distinct TFs.

Table 1. ChIP-seq datasets classified by DBD structure of target TFs

TF classes	Total count of		
	Datasets	Families	TFs
Basic leucine zipper factors (bZIP) {1.1}	96	6	16
Basic helix-loop-helix factors (bHLH) {1.2}	113	6	18
Nuclear receptors with C4 zinc fingers {2.1}	187	3	13
C2H2 zinc finger factors {2.3}	244	3	12
Fork head / winged helix factors {3.3}	92	2	6
Tryptophan cluster factors {3.5}	72	3	13

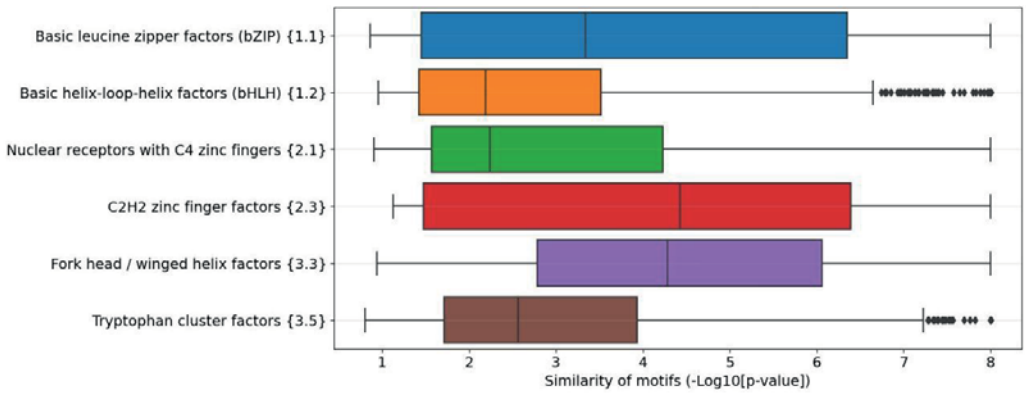


Fig. 1. Intra-class variability of motifs similarity. X-axis denotes classes. Y-axis indicates the significance of motifs similarity, –Log10(p-value). The boxplots depict the distributions of the Q1, Q2 and Q3 quartiles of the fractions of pairs of datasets with certain values of the significance. Whiskers on either side of the Q1/Q3 respect the minimum/maximum values if they were located within 1.5 interquartile ranges (IQR=Q3–Q1) from Q1/Q3, otherwise they are equal to {Q1−1.5*IQR}/{Q3+1.5*IQR}, respectively. In the latter case, we marked all other points as outliers

Results: We applied STREME and TomTom tools to define the enriched motifs of target TFs for all ChIP-seq datasets as described above. Figure 1 shows the distributions of motifs similarity for pairs of enriched motifs within each of six classes. The classes bHLH {1.2} and Nuclear receptors {2.1} known for TF dimerization [7] show the lowest significances (p > 2E-3), while C2H2 zinc finger factors {2.3} and Fork head / winged helix factors {3.3} classes lacking this functionality of TFs show the highest significances (p < 1E-4). Figure 2 compares for six classes the distributions of the median significance of motifs similarity within its families with the cross-family ones. Some pairs of families, e.g. {1.1.1} vs. {1.1.2}, {1.2.1} vs. {1.2.2}, show very high similarity, while the majority of the rest families are not quite similar to other families of the same class, e.g. {1.1.8}, {1.2.6}, and {3.5.1}. We confirmed the structural heterogeneity of TFs from C2H2 zinc finger factors {2.3} class. Only some families due to structural variation of motifs are heterogeneous, e.g. the diagonal cell for the family {1.1.2}.

Conclusion: We propose to indicate for enriched motifs in the results of *de novo* motif search the appropriate level of the hierarchical classification of TFs by DBD structure. These levels should be preliminary defined so that within each of them the motifs are significantly similar.

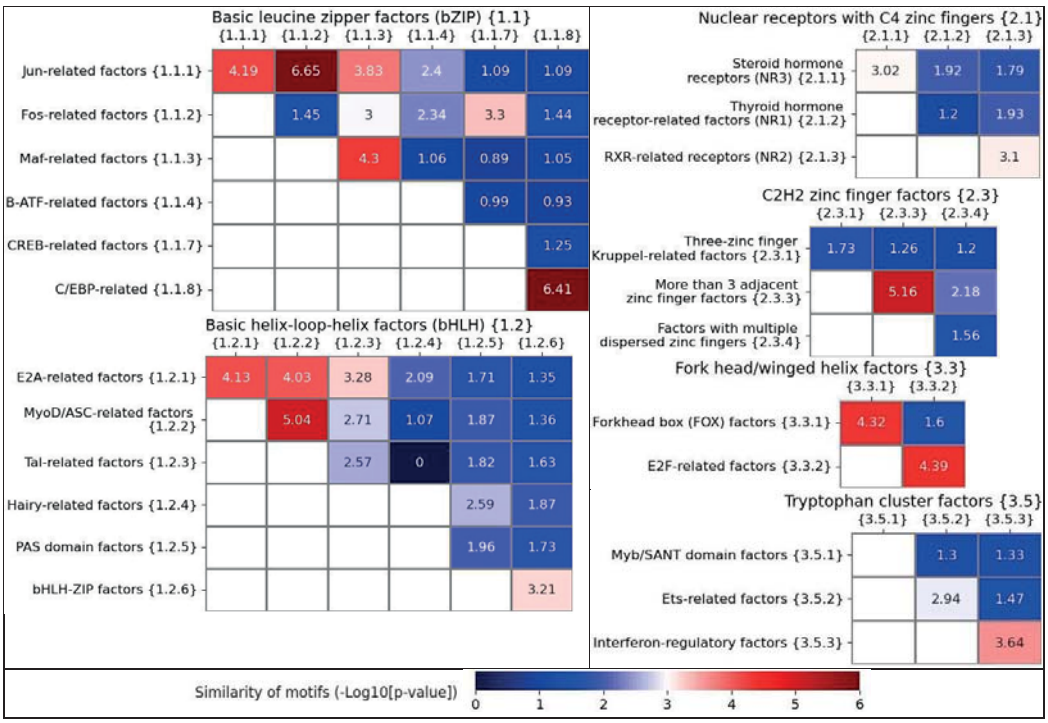


Fig. 2. Variation of the motif similarity within families and between different families for the six classes. Each heatmap entitled with the name of class. Colors/numbers in cells on the diagonals and above them show the median significance for the intra- and inter-family distributions of motifs similarity, –Log10(p-value). Rows and columns show families within the same classes. Empty diagonal cells ({1.1.4}, {1.1.7}, {1.2.4} and {3.5.1}) imply that these families contain sole TFs

Funding: The study is supported Russian Science Foundation project No. 20-14-00140.

References

1. Bailey T.L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics*. 2021;37:2834-2840. doi 10.1093/bioinformatics/btab203
2. Gupta S. et al. Quantifying similarity between motifs. *Genome Biol*. 2007;8(2):R24. doi 10.1186/gb-2007-8-2-r24
3. Rauluseviciute I. et al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2024;52(D1):D174-D182. doi 10.1093/nar/gkad1059
4. Vorontsov I.E. et al. HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Res.* 2024;52(D1):D154-D163. doi 10.1093/nar/gkad1077
5. Wingender E. et al. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* 2018;46:D343-D347. doi 10.1093/nar/gkx987
6. Ambrosini G. et al. Insights gained from a comprehensive all-against-all transcription factor binding motif benchmarking study. *Genome Biol.* 2020;21:114. doi 10.1186/s13059-020-01996-3
7. Amoutzias G.D. et al. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem Sci.* 2008;33(5):220-229. doi 10.1016/j.tibs.2008.02.002
8. Kolmykov S. et al. GTRD: An integrated view of transcription regulation. *Nucleic Acids Res.* 2021;49(D1):D104-D111. doi 10.1093/nar/gkaa1057

doi 10.18699/bgrs2024-1.2-21

New bioinformatics pipeline revealed thousands of novel potential non-canonical binding sites for TDMD-regulated miRNAs in *Drosophila*

Snitkin D.V.[1, 2], Akulenko N.V.[1], Mikhaleva E.A.[1], Ryazansky S.S.[1]*

[1] *National Research Centre "Kurchatov Institute", Moscow, Russia*

[2] *National Research University Higher School of Economics, Moscow, Russia*

* *s.ryazansky@gmail.com*

Key words: miRNA; TDMD; chimeric eCLIP-seq; miRNA binding sites; CLAHer pipeline

Motivation and Aim: During the miRNA-mediated gene silencing in animals, the binding of miRNAs with complementary mRNA targets through a 'seed' region at the 5' end of miRNAs leads to the destruction or translational inhibition of the targets [1]. However, recent high-throughput studies demonstrated that miRNAs mostly prefer to bind with RNA in a non-canonical way, in which the complementary pairings are also located outside the seed region. The biological functions of non-canonical interactions are mostly unknown. The exception is the so-called TDMD (target-directed miRNA degradation) sites, in which miRNA pairs with a target along its whole length with a bulge of several unpaired nucleotides in the middle. Upon binding of miRNAs associated with Ago protein with TDMD sites on RNA targets, Ago is recognized and ubiquitinated by the multiprotein ubiquitin-ligase complex containing substrate receptor ZSWIM8 [2, 3] in mammalian or Dora in *Drosophila* [2, 4]. As a result, Ago undergoes proteolysis, while released miRNA is degraded in the cytoplasm. Although the main features of TDMD sites' structure are known for mammals, their peculiarities in other species remain poorly understood. In our work, we aimed to identify putative TDMD miRNA binding sites with subsequent analysis of their primary and secondary structures in *Drosophila melanogaster* cell culture. The identification and the detailed characterization of features of the *Drosophila* TDMD sites expands our understanding of miRNA biology in general.

Methods and Algorithms: For the identification of the miRNA binding sites on the cognate *Drosophila* RNA targets in OSC cell culture, we used the chimeric eCLIP method combined with PCR enrichment [5] of several TDMD-regulated miRNAs [4]. Additionally, we prepared the eCLIP libraries from cells with depleted Dora. We suggested that TDMD targets would accumulate in Dora-depleted cells thus facilitating their identification. After the high-throughput sequencing, we get ~126M chimeric raw reads in total. For the analysis of the sequenced chimeric molecules, we have developed the bioinformatic pipeline on Python called CLASHer. This pipeline allows pre-processing of the raw reads, identifies the miRNA interaction sites on *Drosophila* transcripts, and reveals various structural features of miRNA binding sites. These features include, for example, the length of the complementary regions between miRNAs and target, the degree of site evolutionary conservation, the site accessibility level, and many others. Finally, we have applied machine learning to classify the non-canonical sites with a focus on the identification of TDMD binding sites.

Results: We identified ~220k chimeras for 5.6k binding sites of 17 miRNAs subjected to TDMD in *Drosophila* OSC cell culture. While the majority (37.1 %) of the identified miRNA binding sites are located in CDS regions of protein-coding mRNAs, only 19.6 % positioned in 3'UTR where the most animal canonical miRNA binding sites are located (Fig. 1A). Only 3.2 % of the binding sites have the seed region while the vast majority of sites (88.6 %) are non-canonical ones without typical seed complementary region (Fig. 1B). We also identified in our data several TDMD targets (e.g. AGO1/miR-999) that were revealed previously [6]; this clearly shows the overall validity of our approach for the identification of the TDMD binding sites.

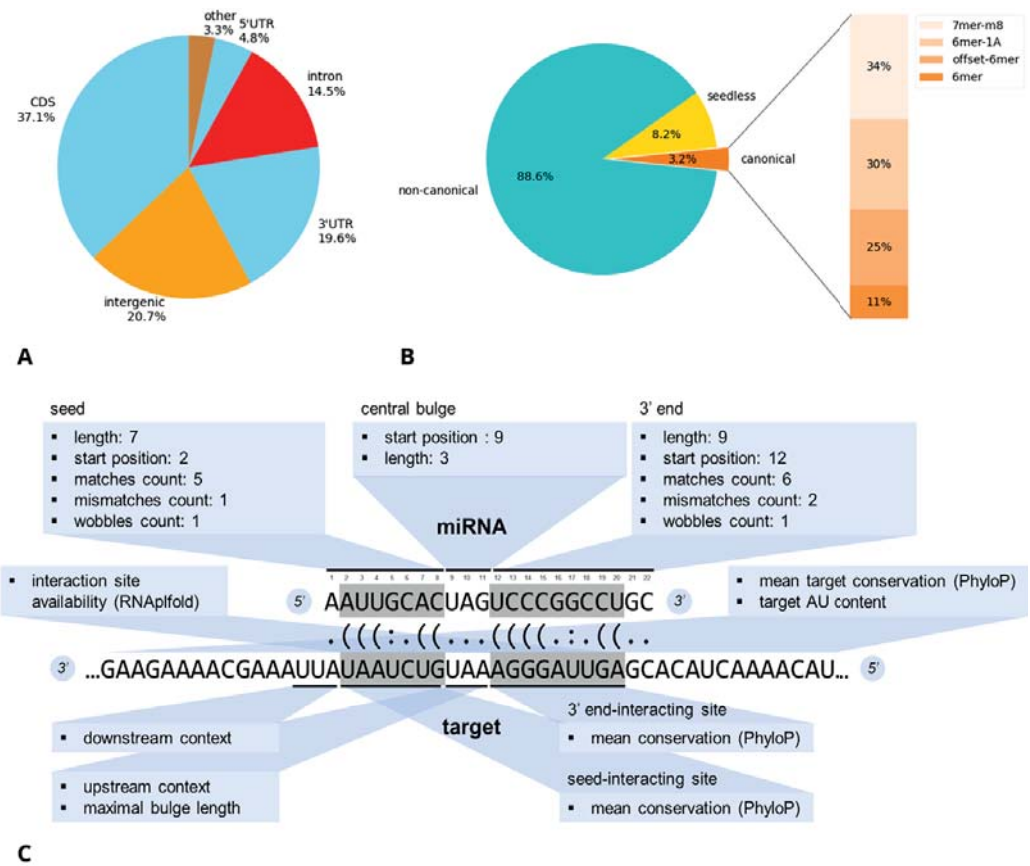


Fig. 1. A, The location of the identified miRNA binding sites. B, The types of the identified miRNA binding sites. C, Some features of the miRNA-target interaction obtained by the pipeline

Then we characterized each binding site by twenty parameters (features), some of them are depicted in Fig. 1C. The correlational and dimensionality reduction methods revealed which features are redundant and should be excluded from subsequent analysis. Finally, we selected the set of ~2k binding sites supported at least by three chimeric molecules in our dataset and classified them based on their features into several groups using machine learning methods. We will present details and results of the work, as well as

attempt to create a predictive algorithm based on machine learning techniques for the identification of TDMD sites.

Conclusion: The developed pipeline allowed to detect 5.6k potential binding sites for TDMD-regulated miRNAs of Drosophila. The conducted detailed characterization of their structural features can be used in building a model for predicting of TDMD binding sites of miRNAs.

Funding: The work was carried out within the state assignment of NRC "Kurchatov Institute".

References

1. Bartel D.P. Metazoan MicroRNAs. *Cell*. 2018;173(1):20-51. doi 10.1016/j.cell.2018.03.006
2. Shi C.Y. et al. The ZSWIM8 ubiquitin ligase mediates target-directed microRNA degradation. *Science*. 2020;370(6523):eabc9359
3. Han J. et al. A ubiquitin ligase mediates target-directed microRNA decay independently of tailing and trimming. *Science*. 2020;370(6523):eabc9546
4. Akulenko N. et al. Evidence of target-mediated miRNA degradation in Drosophila ovarian cell culture. *bioRxiv*. 2023. doi 10.1101/2023.08.30.555489
5. Manakov S.A. et al. Scalable and deep profiling of mRNA targets for individual microRNAs with chimeric eCLIP. *bioRxiv*. 2022. doi 10.1101/2022.02.13.480296
6. Sheng P. et al. Screening of Drosophila microRNA-degradation sequences reveals Argonaute1 mRNA's role in regulating miR-999. *Nat Commun*. 2023;14:2108

doi 10.18699/bgrs2024-1.2-22

Discovery of promising functional motifs as key candidates for multi-antiviral therapy

Sukhanova X.V.[1, 2]*, Huber R.G.[2]

[1] *Applied Genomics Laboratory, SCAMT Institute, ITMO University, St. Petersburg, Russia*

[2] *Biomolecular Function Discovery, Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Matrix #07-01, Singapore, Singapore*

* *sukhanovaxenia@gmail.com*

Key words: RNA interactions; RNA viruses; multiple antiviral therapy; orthogroups; global multiple sequence alignment

Motivation and Aim: In the realm of infectious diseases, the rise of diverse viruses like Zika, Dengue, Enteroviruses (EV-C2, EV71, CA16), SARS-CoV-2, and Chikungunya underscores the need for adaptable antiviral strategies. Each virus presents unique challenges, from severe birth defects associated with Zika and Dengue [1–3] to the global impact of SARS-CoV-2 [6, 7] and the prolonged symptoms caused by Chikungunya [4, 5].

Current antiviral approaches often target specific viruses, leaving a gap in developing interventions effective across various infections [8–10]. Recent advances in understanding RNA-RNA interactions have prompted exploration into shared intervention regions across different viruses. The SPLASH method, known for its comprehensive study of RNA-RNA interactions, offers a practical avenue for this exploration.

Methods and Algorithms: In our pursuit of adaptable antiviral strategies, we employed a multifaceted methodology to identify potential targets for versatile antiviral therapy applicable across various viral infections (Fig. 1). The search began by assigning orthogroups to cluster host transcripts, aiming to identify shared orthogroups across at least five viruses. This step allowed to delineate transcript clusters essential for subsequent analysis. Subsequently, we delved into functional annotation and molecular network interactions using the STRING database. This phase involved annotating host transcripts within orthogroups and exploring their interactions within cellular networks, providing insights into potential intervention regions. To identify significant viral RNA motifs with multi-viral therapeutic potential, we employed advanced sequence analysis techniques. This included GC content Markov Chain Monte Carlo simulations and multiple-sequence global alignment, enabling us to detect highly significant RNA motifs amidst the genomic complexity. Following motif identification, highly probable motifs were estimated by using position-specific score matrix and genome scanning. This step allowed us to pinpoint motifs with the highest likelihood of exhibiting broad-spectrum efficacy against diverse viral infections.

Results: We identified orthogroup-specific viral motifs, notably OG0002480 and OG0009649, exhibiting promising characteristics for broad-spectrum antiviral interventions. These motifs demonstrated specificity across diverse viral genomes, including Dengue, SARS-CoV2 Delta, and Zika (Brazilian strain) viruses for OG0002480 motif, and Zika (Brazilian, Singaporean strains) and SARS-CoV2 (wt and Delta strains) viruses for OG0009649 motif. Pairwise intersections between unrelated

viruses and genome scanning allowed us to identify regions of high confidence in RNA interactions. With strain-specific genomic origin, highlighting the nuanced diversity in viral infection mechanisms, even among closely related viruses. Exploration into the cellular mechanisms associated with these motifs provided crucial insights into viral infection dynamics, particularly highlighting their involvement in actin binding and focal adhesion through ACTN1, ACTN4, VCL and FLNC1 FLNC3 transcripts, consequently.

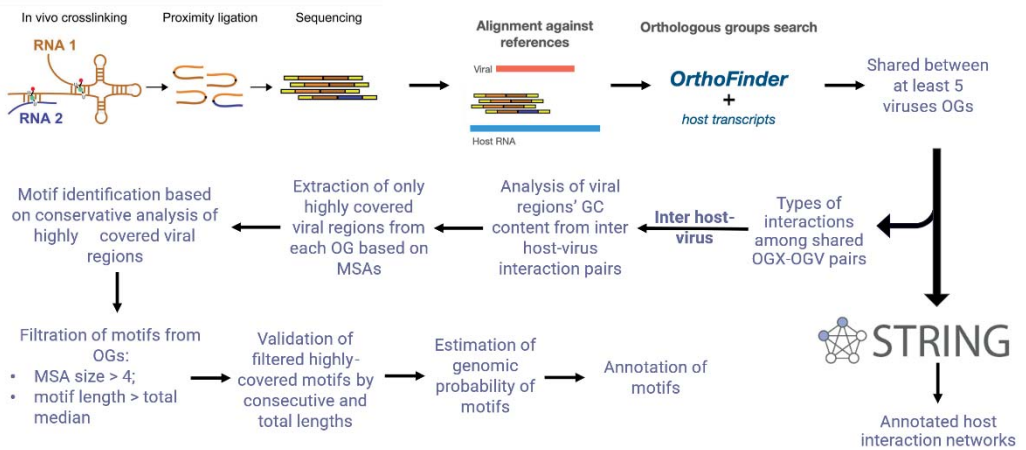


Fig. 1. Pipeline of the identification and analysis of potential motifs as multiple antiviral therapy targets. Step 1. SPLASH data analysis. Step 2. Orthogroups assignment and search for conservative ones. Step 3. Inter-host-viral interactions filtration and potential motifs identification. Step 4. Validation of identified motifs via genome scanning. Step 5. Annotation of identified motifs and host interaction network analysis

Conclusion: The applied analysis of SPLASH interactions between viral and host RNA laid a robust foundation for antiviral therapy development. The identification of orthogroup-specific motifs underscores promising avenues for therapeutic intervention, potentially offering broad-spectrum effectiveness. As we continue to unravel the intricacies of viral-host interactions, the prospects for developing effective, multi-viral therapies appear increasingly promising. Moving forward, further exploration of shared RNA motifs and validation of potential therapeutic targets will be crucial for advancing the field of antiviral therapy development and combating emerging viral threats effectively.

References

1. de Araújo T.V.B. et al. Association between microcephaly, Zika virus infection, and other risk factors in Brazil: final report of a case-control study. *Lancet Infect Dis*. 2018;18(3):328-336. doi 10.1016/S1473-3099(17)30727-2
2. Krauer F. et al. Zika Virus Infection as a Cause of Congenital Brain Abnormalities and Guillain–Barré Syndrome: Systematic Review. *PLoS Med*. 2017;14(1):e1002203. doi 10.1371/journal.pmed.1002203
3. Bhatt S. et al. The global distribution and burden of dengue. *Nature*. 2013;496(7446):504-507. doi 10.1038/nature12060

4. Cabrerizo M. et al. Molecular epidemiology of enterovirus 71, coxsackievirus A16 and A6 associated with hand, foot and mouth disease in Spain. *Clin Microbiol Infect*. 2014;20(3):O150-O156. doi 10.1111/1469-0691.12361
5. Solomon T., Lewthwaite P., Perera D., Cardosa M.J., McMinn P., Ooi M.H. Virology, epidemiology, pathogenesis, and control of enterovirus 71. *Lancet Infect Dis*. 2010;10(11):778-790. doi 10.1016/S1473-3099(10)70194-8
6. Zhu N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727-733. doi 10.1056/NEJMoa2001017
7. Silva J.V.J. et al. A scoping review of Chikungunya virus infection: epidemiology, clinical characteristics, viral co-circulation complications, and control. *Acta Trop*. 2018;188:213-224. doi 10.1016/j.actatropica.2018.09.003
8. Poltronieri P., Sun B., Mallardo M. RNA viruses: RNA roles in pathogenesis, coreplication and viral load. *Curr Genomics*. 2015;16(5):327-335. doi 10.2174/1389202916666150707160613
9. Nicholson B.L., White K.A. Functional long-range RNA-RNA interactions in positive-strand RNA viruses. *Nat Rev Microbiol*. 2014;12(7):493-504. doi 10.1038/nrmicro3288
10. Sola I., Mateos-Gomez P.A., Almazan F., Zuñiga S., Enjuanes L. RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol*. 2011;8(2):237-248. doi 10.4161/rna.8.2.14991

doi 10.18699/bgrs2024-1.2-23

Identification of potential transcription factors from whole transcriptome data by taking into account enrichment and classification of motifs of known transcription factors, and their expression levels

Tsukanov A.V.*, Levitsky V.G.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** tsukanov@bionet.nsc.ru*

Key words: transcription factors; motif enrichment; RNA-seq; transcription; regulation of transcription

Motivation and Aim: To date, RNA-seq remains the most popular technology for the whole transcriptome analysis identifying differentially expressed genes (DEGs). An important challenge in the analysis of RNA-seq data is the search for transcription factors (TFs) associated with changes in gene expression levels in response to external stimuli. A straightforward approach to solve this issue is to define the enriched motifs of known TFs in promoters of DEGs. We previously proposed an ESDEG tool (https://github.com/ubercomrade/esdeg) [1] that implemented this approach, but its efficiency may be further improved. While the number of known motifs is approached the number of known TFs (above 1400 and 1600 for human [2, 3]), the number of structurally different motifs is substantially smaller [4], and structurally-related TFs often have very similar motifs. Hence, a number of distinct enriched motifs can reach several hundreds, and each of them often respects at least several TFs. Thus, now we are still too far from a reliable definition of key TFs associated with the change of the whole transcriptome. Here we apply information on the classification of TFs [5] and their expression levels to refine the lists of reliable TFs from RNA-seq data.
Methods and Algorithms: As input data, we use a counting matrix in which genes are rows, the test and control samples are columns, and cells are the numbers of reads mapped to genes. Next, the *pyDEseq2* package [6] determines DEGs (we apply the criteria $|log_2(Fold\ Change)| > 1$ and $padj < 0.05$), the *rnanorm* tool (https://github.com/genialis/RNAnorm) calculates *FPKM* values for genes (normalized read counts based on gene length and the total number of mapped reads). Further, the ESDEG tool is applied. It takes an annotation of DEGs from *pyDEseq2* and the list of all motifs as positional frequencies matrices from the HOCOMOCO database [2]. The ESDEG identifies the list of enriched motifs in promoters of DEGs using a Monte Carlo approach. For each motif the enrichment (odds) ratio ($log_2(OR)$) and significance (*em_padj*) are defined as described earlier [1]. Here we introduced a few changes to ESDEG as follows. For each motif, class and family information for TFs [2, 5] is provided. Then, the expression level values (*FPKM*) as well as expression level changes ($log_2(Fold\ Change)$) and theirs significances (*padj*) are given for TFs of all motifs. As a result, a common table combines expression levels and motif enrichment information for each TF. The final list of TFs is determined as follows. First, we restrict the most significant enriched motifs, $em\_padj < 0.05$ and $log_2(OR) > 1.0$. Second, we test whether the most reliable TF genes are found among DEGs. Third, only those TFs with *FPKM*

values above a specified threshold are left. The threshold was set as the third quartile (Q_3) of the *FPKM* distribution over all TFs.

Results: The count matrix for the RNA-seq data of the E-MTAB-6598 experiment (human K562 cell line, interferon alpha treatment) we obtained from the Expression Atlas database (https://www.ebi.ac.uk/gxa/home). We used 5'-regions (–400; +100) relative gene transcription start sites in analysis. The *pyDEseq2* determined 159 DEGs, and ESDEG found totally 226 significantly enriched motifs for 173 TFs among 1443 motifs for 949 TFs. They were classified into 42 families [2, 5] (Fig. 1).

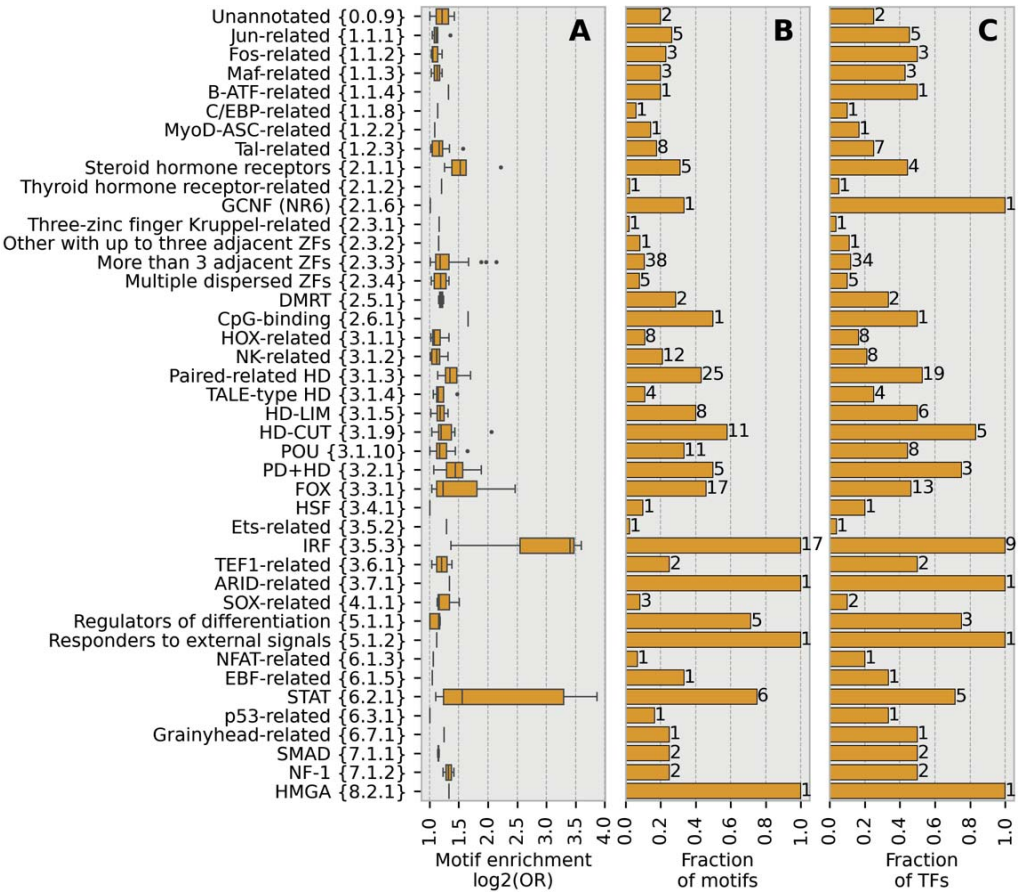


Fig. 1. Results of ESDEG tool application for the RNA-seq data of the E-MTAB-6598 experiment. **A**, Distribution of the motif enrichment for TF families. **B** and **C**, Fractions of enriched motifs and their TFs per TF family, for each family the numbers of enriched motifs and TFs are marked

The highest enrichment we found for the motifs of TFs from the IRF {3.5.3} and STAT {6.2.1} families (Fig. 1A). However, the fractions of enriched motifs notably varied between different families (Fig. 1B). Motifs of all nine TFs from the IRF family are enriched, five out of total seven TFs from the STAT family have enriched motifs (71 %, Fig. 1C). TFs from the IRF and STAT families are involved in signaling pathways related to the regulation of type I interferon [7]. Unfortunately, the results of motif enrichment analysis are insufficient to identify specific TFs exactly. The TF-DEG matching criterion left only four most reliable TFs from three families (Table 1, bold).

The criterion on the gene expression level gave a few families and TFs with enriched motifs. Here we kept only TFs with FPKM > 4.3 (the value of Q_3). This allowed us to predict 31 additional TFs from 20 TF families. Table 1 shows TFs from 12 families with the highest enrichment of motifs. Totally, among 173 TFs predicted by ESDEG tool, the TF-DEG matching and gene expression level criteria mark 35 TFs as reliable.

Table 1. Application of the gene expression level and TF-DEG matching criteria to the results of ESDEG analysis. Bold font marks DEGs representing TFs

TF family	Number of TFs before / after filtration		List of TFs
Maf-related {1.1.3}	3	1	MAFG
Tal-related {1.2.3}	7	1	TAL1
Steroid hormone receptors {2.1.1}	4	1	NR3C1
Thyroid hormone receptor-related {2.1.2}	1	1	RARA
Three-zinc finger Kruppel-related {2.3.1}	1	1	KLF3
Paired-related HD {3.1.3}	18	1	**HESX1**
FOX {3.3.1}	13	1	FOXM1
IRF {3.5.3}	9	3	**IRF9**, **IRF7**, IRF3
TEF1-related {3.6.1}	2	2	TEAD3, TEAD1
ARID-related {3.7.1}	1	1	ARID1A
STAT {6.2.1}	5	4	**STAT1**, STAT6, STAT5B, STAT2
HMGA {8.2.1}	1	1	HMGA1

Conclusion: We have developed an approach that applied additional information on TF classification and their expression levels to deduce TF regulators from RNA-seq data. Our approach allowed us to identify the most reliable TFs involved in DEG regulation among a large number of TFs whose motifs can be enriched in DEG promoters.

Funding: The study is supported by Russian Science Foundation project No. 20-14-00140.

References

1. Oshchepkov D. et al. Transcription factors as important regulators of changes in behavior through domestication of gray rats: quantitative data from RNA sequencing. *Int J Mol Sci*. 2022;23(20):12269. doi 10.3390/ijms232012269
2. Vorontsov I.E. et al. HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Res.* 2024;52(D1):D154-D163. doi 10.1093/nar/gkad1077
3. Lambert S.A. et al. The human transcription factors. *Cell*. 2018;172(4):650-665. doi 10.1016/j.cell.2018.01.029
4. Ambrosini G. et al. Insights gained from a comprehensive all-against-all transcription factor binding motif benchmarking study. *Genome Biol.* 2020;21(1):114. doi 10.1186/s13059-020-01996-3
5. Wingender E. et al. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* 2018;46(D1):D343-D347. doi 10.1093/nar/gkx987
6. Muzellec B. et al. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics.* 2023;39(9):btad547. doi 10.1093/bioinformatics/btad547
7. Mogensen T.H. IRF and STAT transcription factors – from basic biology to roles in infection, protective immunity, and primary immunodeficiencies. *Front Immunol.* 2019;9:3047. doi 10.3389/fimmu.2018.03047

doi 10.18699/bgrs2024-1.2-24

Motif discovery and benchmarking from diverse assays highlight the best tools for predicting DNA-binding specificity of human transcription factors

Vorontsov I.E.[1], Kozin I.[2, 3], Abramov S.[4], Boytsov A.[4], Jolma A.[5], Albu M.[5], Ambrosini G.[6, 7], Faltejskova K.[8], Gralak A.[6, 7], Gryzunov N.[9], Inukai S.[10], Kolmykov S.[11], Kravchenko P.[12], Kribelbauer J.[6, 7], Laverty K.[5], Nozdrin V.[9], Patel Z.[5], Penzar D.[1], Plescher M.-L.[13], Pour S.[5], Razavi R.[5], Yang A.[5], Yevshin I.[11], Zinkevich A.[9], Bucher P.[7], Deplancke B.[6, 7], Fornes O.[14], Grau J.[13], Grosse I.[13], Kolpakov F.[11], GRECO-BIT/Codebook Consortium, Makeev V.J.[1], Hughes T.[5], Kulakovskiy I.V.[1, 3]*

[1] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*
[2] *Faculty of Biotechnology, Lomonosov Moscow State University, Moscow, Russia*
[3] *Institute of Protein Research, RAS, Pushchino, Russia*
[4] *Altius Institute for Biomedical Sciences, Seattle, WA, USA*
[5] *Donnelly Centre and Department of Molecular Genetics, Toronto, Canada*
[6] *School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*
[7] *Swiss Institute of Bioinformatics, Lausanne, Switzerland*
[8] *Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Praha, Czech Republic; Computer Science Institute, Faculty of Mathematics and Physics, Charles University, Praha, Czech Republic*
[9] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*
[10] *Chugai Pharmaceutical Co., Ltd, Tokyo, Japan*
[11] *Department of Computational Biology, Sirius University of Science and Technology, Sirius, Krasnodar region, Russia*
[12] *Max Planck Institute of Biochemistry, Planegg, Germany*
[13] *Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle, Germany*
[14] *Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, University of British Columbia, Vancouver, BC, Canada*
* *ivan.kulakovskiy@gmail.com*

Key words: TF; TFBS; DNA motif discovery; benchmarking; transcription factors

Motivation and Aim: Computational methods that discover DNA motifs specifically binding proteins have been developed for over three decades, stimulated by the continuing progress in experimental methods for DNA-protein interaction profiling. Yet, the assessment of how well the computational models quantitatively reflect the binding specificities is still far from routine. Particularly, despite the availability of many advanced motif models, the position weight matrix, the traditional model of DNA binding specificity, remains the most commonly used approach, and it is still unclear if the more complex model outperforms positional weight matrices, and if so, to which degree.
Methods: As GRECO-BIT/Codebook initiative, we used the newly generated experimental data from several thousand experiments for several hundred less-studied

human transcription factors to derive binding motifs with diverse DNA motif discovery tools, from multiple sequence aligners to deep learning methods. First, we performed an extensive benchmarking study highlighting the best-performing tools and forming a rich catalog of motifs, many of which belong to previously unexplored transcription factors. Next, we used a carefully selected subset of forty TFs to call for the wisdom of crowds: we have organized an open competition in constructing the best sequence-level models of transcription factor binding sites from these yet unpublished data. The challenge in Inferring Binding Specificities, IBIS, is built on the experimental data on the TF-DNA interactions obtained from five different experimental techniques, and aimed at a fair assessment of existing and novel methods for de novo TFBS motif discovery, using both classic Position Weight Matrices and Arbitrary Advanced Approaches, "triple-A" models.

Results: We plan to report the best-performing software tools and key findings of the internal within-Consortium motif benchmarking study and tease the first insights from the open IBIS Challenge, which will end mid-summer 2024.

Funding: The study is supported by RSF 20-74-10075 to IVK.

doi 10.18699/bgrs2024-1.2-25

Unlocking DNA methylation with BSXplorer

Yuditskiy K.I., Bezdvornykh I.V., Samsonova A.A., Kanapin A.A.*

Institute of Translational Biomedicine, Saint Petersburg State University, St. Petersburg, Russia

** akanapin@spbu.ru*

Key words: epigenomics; bisulphite sequencing; exploratory analysis; transcription regulation; methylation density

Motivation and Aim: Cytosine methylation is a common epigenetic mark that is generally found in eukaryotes, including vertebrates, insects, fungi, and plants, exhibiting distinct patterns and pathways of DNA methylation [1, 2]. Bisulfite sequencing detects and quantifies DNA methylation patterns, contributing to our understanding of gene and transposon expression regulation, genome stability maintenance, conservation of epigenetic mechanisms. Despite being a technology of choice to profile DNA methylation for many years now there are surprisingly few lightweight and robust standalone tools available for efficient graphical analysis of data in non-model systems. Graphical representation of methylation data is crucial in exploring epigenetic regulation on a genome-wide scale however this feature is missing in many of existing analytical packages [3, 4]. This is especially relevant for non-model organisms with poorly annotated genomes and/or organisms where genome sequences are not yet assembled on chromosome level. BSXplorer is a tool specifically developed to assist researchers in explorative data analysis and in visualising and interpreting bisulfite sequencing data more easily.

Methods and Algorithms: BSXplorer is implemented in Pyhton (version 3.9 or higher). The package runs on most modern systems and its functions are available through both Python API and command-line interface. The data processing speed is mostly limited by I/O capacity of the storage. BSXplorer is publicly available at https://github.com/shitohana/BSXplorer or https://pypi.org/project/bsxplorer/ where a comprehensive user manual is provided. The tool processes methylation data quickly and offers API and CLI capabilities, along with the ability to create high-quality figures suitable for publication.

Results: BSXplorer provides in-depth graphical and exploratory analysis of bisulphite sequencing data and includes the following features: profiling of methylation levels in metagenes or in user-defined regions, generation of summary statistics charts, comparative analyses of methylation patterns across experimental samples, methylation contexts and species, and, finally, identification of modules sharing similar methylation signatures in various functional genomic elements.

Conclusions: Overall, BSXplorer is a lightweight and flexible instrument that facilitates explorative analyses of DNA methylation patterns, including BS-seq data mining, contrasting and visualization of methylation patterns in genomes of model and non-model organisms. The tool can be used on its own or seamlessly incorporated into epigenomic data processing workflows.

Funding: The study is supported by Russian Scientific Foundation (Grant No. 23-14-00134).

References

1. Greenberg M.V.C., Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol*. 2019;20(10):590-607. doi 10.1038/s41580-019-0159-6
2. Zhang H., Lang Z., Zhu J. K. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Boil.* 2018;19(8):489-506. doi 10.1038/s41580-018-0016-z
3. Müller F., Scherer M., Assenov Y., Lutsik P., Walter J., Lengauer T., Bock C. RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biol*. 2019;20(1):55. doi 10.1186/s13059-019-1664-9
4. Ki H., Si M., Par N., Kwo K., Ki J., Ki J. msPIPE: a pipeline for the analysis and visualization of whole-genome bisulfite sequencing data. *BMC Bioinformatics*. 2022;23(1):383. doi 10.1186/s12859-022-04925-2

doi 10.18699/bgrs2024-1.2-26

Comprehensive mapping and modelling of the rice regulome landscape unveils the regulatory architecture underlying complex traits

Zhu T.[1], Xia C.J.[2], Yu R.R.[1], Zhou X.K.[1], Xu X.B.[2], Wang L.[1], Zong Z.X.[2], Yang J.J.[2], Liu Y.M.[2], Ming L.C.[2], You Y.X.[1], Chen D.J.[1]*, Xie W.B.[2]**

[1] *State Key Laboratory of Pharmaceutical Biotechnology, Department of Gastroenterology, Nanjing Drum Tower Hospital, National Resource Center for Mutant Mice, School of Life Sciences, Nanjing University, Nanjing, China*

[2] *National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan, China*

** dijunchen@nju.edu.cn (Dijun Chen); ** weibo.xie@mail.hzau.edu.cn (Weibo Xie)*

Key words: rice; ATAC-seq; RNA-seq; GWAS; deep learning

Motivation and Aim: Unraveling the regulatory mechanisms that govern complex traits is pivotal for advancing crop improvement [1]. However, our understanding of the regulatory mechanisms governing complex traits in rice remains incomplete. Given that diverse traits manifest across distinct developmental stages and tissues, systematic annotation of noncoding regulatory variants in rice is currently hindered by the lack of a comprehensive epigenome map across various tissues and growth stages.

Methods and Algorithms: UMI-ATAC-seq [2] was used to capture chromatin open regions in rice. cisDynet [3] was used for data preprocessing and downstream advanced analysis of ATAC-seq. RNA-seq data were quantified using RSEM [4]. Basenji [5] deep learning model was used to train with prediction of tissue-specific chromatin accessibility and used to assess the effect of non-coding region variants.

Results: we systematically mapped chromatin accessibility profiles in various tissues across the life cycle of three representative rice cultivars using the UMI-ATAC-seq method [2], a modified ATAC-seq (assay for transposase accessible chromatin-sequencing) protocol developed in our lab. Through analysis of 145 ATAC-seq datasets, we obtained a total of 117,176 unique open chromatin regions (OCRs), accounting for ~15 % of the rice genome. By integration of RNA-seq data from matched tissues, we predicted potential target genes for OCRs based on the correlation of gene expression and adjacent chromatin accessibility across tissues. Through TF footprinting analysis, we inferred tissue- or stage-specific regulatory networks and identified cultivar-polymorphic/trait-associated OCRs by comparing the regulatory landscapes between indica and japonica rice subspecies. Notably, our analysis unveiled a preference for GWAS-associated variants within tissue-specific OCRs, enabling the identification of causal associations between 209 complex agronomic traits and noncoding regulatory variants using this OCR landscape. Utilizing optimized deep learning models, we decoded the regulatory grammar through modeling of tissue-specific chromatin accessibility and across-variety predictions from sequences. The modeling approach sheds light on the key genetic alterations contributing to cis-regulatory divergence. Overall, these data not only serve as a cornerstone resource for the plant research

community but also provide valuable regulatory variants for precision molecular breeding.

Conclusion: We used UMI-ATAC-seq to map chromatin accessibility in three rice varieties. Combined with RNA-seq, we effectively established associations between regulatory regions and target genes. Using these resources, we integrated published GWAS data and cloned the gene *Osbzip06*, which affects seed germination rate. We optimized deep learning models that can effectively predict tissue-specific chromatin accessibility and help to identify important genetic loci associated with agronomic traits. Overall, our study establishes a foundational resource for rice functional genomics and precision molecular breeding, providing valuable insights into regulatory mechanisms governing complex traits.

Funding: This study was supported by the National Natural Sciences Foundation of China (No. 32070656, 31922065).

References

1. Fu L.-Y. et al. ChIP-Hub provides an integrative platform for exploring plant regulome. *Nat Commun*. 2022;13(1):3413
2. Zhu T. et al. ATAC-seq with unique molecular identifiers improves quantification and footprinting. *Commun Biol*. 2020;3(1):675
3. Zhu T. et al. cisDynet: An integrated platform for modeling gene-regulatory dynamics and networks. *iMeta*. 2023;2(4):e152
4. Li B., Dewey C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12(1):323
5. Kelley D.R. et al. Sequential regulatory activity prediction across chromosomes with convolutional neural networks. *Genome Res*. 2018;28(5):739-750

doi 10.18699/bgrs2024-1.2-27

PARADE: deep learning for prediction and rational design of untranslated regions with desired tissue-specific expression

Zinkevich A.[1]*, Aristova E.[1], Lee S.[2], Mittmann T.[2], Penzar D.[3, 4], Yousefi H.[2], Khoroshkin M.[2], Kulakovskiy I.V.[4]

[1] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

[2] *University of California, San Francisco, CA, USA*

[3] *Artificial Intelligence Research Institute (AIRI), Moscow, Russia*

[4] *Institute of Protein Research, RAS, Pushchino, Russia*

* *arsenii.z.work+bgrs@gmail.com*

Key words: synthetic biology; diffusion neural network; generative modelling; rational design; regulatory genomics, RNA-regulation; artificial intelligence

Motivation and Aim: Gene expression is a multi-level process that determines the morphological and functional characteristics of a cell. Different levels of gene expression are controlled by diverse regulatory mechanisms, from DNA epigenetic marks to post-translational modification of proteins. A significant regulatory layer is mediated by mRNA untranslated regions (5'UTR and 3'UTR), which significantly influence the mRNA translation and stability of the transcribed RNA molecules.

With the advent and rapid development of massively parallel reporter assays, it became possible to profile the regulatory effects of thousands of UTRs and identify and model the sequence-level regulatory grammar with the help of deep learning. Yet, existing studies fail to properly capture and predict the UTR-dependent pattern of cell type-specific mRNA translation and stability.

Methods and Algorithms: This study uses the results of MPRAs [1] performed with 20k+20k 5' and 3' UTR sequences in five distinct cell types. To analyze the data, we adapted the LegNet neural network [2] to the regression of the normalized reporter fluorescence from the UTR sequence in the particular cell type. The neural network was implemented in Python within the PyTorch framework.

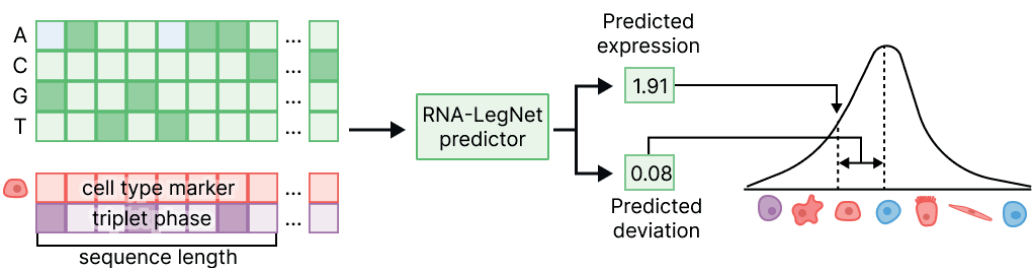


Fig. 1. Simplified double regression task. A neural network model is trained to simultaneously predict UTR sequence expression and expression difference from the mean

Results and Conclusions: Here we present the PARADE (Prediction and rational design of RNA UTRs) framework, a deep learning solution for estimating cell type-specific activity of 5'- and 3'UTRs, and performing rational design of UTR sequences constrained by desired expression patterns. To generate new sequences, PARADE employs 3 different methods: fine-grain filtering of random sequences, genetic algorithm, and diffusion-denoising neural network. The performance of the methods was evaluated and validated experimentally.

In the end, PARADE predicts the expression levels for the test set with r = 0.824 for 5'UTR and r = 0.793 for 3'UTR sequences, outperforming the previous specialized Optimus-5-prime model [3], and successfully generates sequences with desired directional tissue-specific activity.

Funding: A.Z. was supported by a personal fellowship from the Non-commercial Foundation for Support of Science and Education "INTELLECT". D.P. and I.V.K. were supported by RSF grant 20-74-10075.

References

1. Litterman A.J., Kageyama R., Le Tonqueze O. et al. A massively parallel 3' UTR reporter assay reveals relationships between nucleotide content, sequence conservation, and mRNA destabilization. *Genome Res*. 2019;29(6):896-906. doi 10.1101/gr.242552.118
2. Penzar D., Nogina D., Noskova E. et al. LegNet: a best-in-class deep learning model for short DNA regulatory regions. *Bioinformatics*. 2023;39(8):btad457. doi 10.1093/bioinformatics/btad457
3. Sample P.J., Wang B., Reid D.W., Presnyak V., McFadyen I.J., Morris D.R., Seelig G. Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat Biotechnol*. 2019;37(7):803-809. doi 10.1038/s41587-019-0164-5

1 Симпозиум «Геномика, транскриптомика и биоинформатика»

Symposium “Genomics, transcriptomics and bioinformatics”

doi 10.18699/bgrs2024-1.3-01

Изучение влияния дупликации гена *STAG2* в раннем нейральном развитии

Гридина М.[1, 2]*, Хабарова А.[1], Степанчук Я.[1, 2], Нурисламов А.[1, 2], Кокшарова Г.[1], Крепиши А.[3], Коста С.[3], Фишман В.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Университет Сан-Паулу, Сан-Паулу, Бразилия*

* *gridinam@gmail.com*

Ключевые слова: когезин; индуцированные плюрипотентные стволовые клетки; нейральная дифференцировка; когезинопатии

Мотивация и цель: Когезин – эволюционно консервативный мультимерный белковый комплекс, состоящий из нескольких субъединиц: SMC1A, SMC3, RAD21 и STAG1 или STAG2. Когезин – важный белок митотического цикла, обеспечивающий когезию сестринских хроматид, участник процессов репарации двуцепочечных разрывов ДНК по механизму гомологичной рекомбинации [1]; участвует во взаимодействиях между цис-регуляторными элементами генома [2]. Мутации белков когезинового комплекса связаны с возникновением заболеваний, которые получили названия когезинопатии. Наиболее охарактеризованной из них является синдром Корнелии де Ланге, вызываемый мутациями в генах *NIPBL*, *SMC1A* или *SMC3* [3, 4]. Микродупликации Xq25, включающие всю последовательность гена *STAG2*, также вызывают когезинопатии, основными фенотипическими проявлениями которых являются: гипотония, задержка развития речи, умственная отсталость от легкой до умеренной степени и аномальное поведение [5, 6]. Однако механизм развития патологии в результате дупликации *STAG2* неизвестен. Ранее были обнаружены два пациента с амплификацией района Xq25 размером 377 kb (chrX:123056490-123433630, GRCh37), которая затрагивает единственный ген – *STAG2*. Пациенты – сводные братья, у обоих наблюдается задержка развития и аутистические черты. Целью исследования является определение роли нарушений копийности гена *STAG2* на ранних этапах развития клеток мозга. Мы ожидаем, что эти эксперименты прольют свет на молекулярные механизмы развития патологии, лежащие в основе эффекта амплификации Xq25.
Методы и алгоритмы: Для получения индуцированных плюрипотентных стволовых клеток (ИПСК) человека из мононуклеарных клеток периферической крови мы использовали котрансфекцию эписомальных векторов, несущих следующие гены: *OCT4*, *MYC*, *LIN28*, shRNA p53, *SOX2*, *KLF4*, *EBNA1*, – согласно протоколу [7]. Дифференцировку в нейральные клетки предшественники проводили через образование эмбриоидных телец и последующую селекцию нейральных розеток [8]. РНК была выделена с использованием набора Hi pure Total RNA Plus kit (Magen), согласно инструкции производителя. Секвенирование polyA-селектированной РНК проводили на платформе DNBSEQ парными прочтениями по 100 п.н. Поиск дифференциально экспрессирующихся генов проводили при помощи DESeq2.

Результаты: У обоих пациентов наблюдается задержка развития и аутистические черты, поэтому можно предположить, что дупликация гена *STAG2* приводит к нарушению раннего развития головного мозга. Поэтому в качестве целевой популяции клеток для оценки эффектов микродупликации мы решили использовать клетки нейральных предшественников. На первом этапе из мононуклеарных клеток периферической крови мы получили линии ИПСК для обоих пациентов и здоровых доноров. Для каждого генотипа были выбраны по три линии ИПСК. Увеличение копийности генов не всегда приводит к увеличению уровня их экспрессии [8]. Мы показали, что уровень экспрессии *STAG2* был выше во всех проанализированных линиях ИПСК пациентов, по сравнению с клетками здоровых доноров. Далее мы дифференцировали выбранные ИПСК, получили клетки нейральных предшественников и провели анализ транскриптома. PCA по всем генам показал, что реплики хорошо сходятся между собой. Также мы обнаружили ряд особенностей транскрипции в клетках нейральных предшественников пациентов.
Выводы: Микродупликация гена *STAG2* приводит к увеличению его экспрессии в ИПСК, по сравнению с клетками с нормальным кариотипом. В нейральных клетках-предшественниках с дупликацией гена *STAG2* обнаружен ряд особенностей транскрипции по сравнению с клетками с нормальным кариотипом.
Финансирование: Исследование поддержано Российской Федерацией в лице Минобрнауки России, номер Соглашения № 075-15-2024-539 от 24.04.2024.

Influence of *STAG2* gene duplication in early neural development

Gridina M.[1, 2]*, Khabarova A.[1], Stepanchuk Ya.[1, 2], Nurislamov A.[1, 2], Koksharova G.[1], Krepishi A.[3], Costa S.[3], Fishman V.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *University of Sao Paulo, Sao Paulo, Brasil*

* *gridinam@gmail.com*

Key words: cohesin; induced pluripotent stem cells; neuronal differentiation; cohesinopathies

Motivation and Aim: Cohesin is an evolutionarily conserved protein complex consisting of several subunits: SMC1A, SMC3, RAD21 and STAG1 or STAG2. Cohesin is an important protein of the mitotic cycle, involving in the cohesion of sister chromatids, in a the repair of double-stranded DNA breaks through the mechanism of homologous recombination [1] and in interactions between cis-regulatory elements of the genome [2]. Mutations in the cohesin complex proteins are associated with the called cohesinopathies. The best characterized of these is Cornelia de Lange syndrome, caused by mutations in the *NIPBL*, *SMC1A* or *SMC3* genes [3, 4]. Microduplications of Xq25, including the entire *STAG2* gene sequence, also cause cohesinopathies, the main phenotypic manifestations of which are: hypotonia, delayed speech development, mild to moderate mental retardation and abnormal behavior [5, 6]. However, the mechanism of the pathology development is unknown. Previously, two patients were identified with amplification of the Xq25 region of 377 kb (chrX:123056490-123433630, GRCh37),

which affects a single gene, *STAG2*. The patients are half-brothers and both have developmental delays and autistic traits. The purpose of the study is to determine the role of *STAG2* gene copy number variations in the early stages of brain cell development. We expect that these experiments will shed light on the molecular mechanisms of diseases resulting from Xq25 amplification.

Methods and Algorithms: Reprogramming of human blood cells into induced pluripotent stem (iPS) cells was performed by episomal vectors cocktail contained: reprogramming vectors expressed OCT4, MYC, LIN28, shRNA against p53, SOX2, KLF4 and EBNA1, according to the previously described protocol [7]. iPS cells were cultured and differentiated into neural progenitor cells as described before [8]. RNA was isolated using the Hi pure Total RNA Plus kit (Magen) according to the manufacturer's instructions. Sequencing of polyA-selected RNA was performed on the DNBSEQ platform using paired 100 bp reads. The statistical analysis and identification of differentially expressed genes was performed using DEseq2.

Results: Both patients have developmental delay and autistic features, suggesting that *STAG2* gene duplication leads to impairment of early brain development. Therefore, we decided to use neural progenitor cells as the target cell population to evaluate the effects of microduplication. We obtained iPS cell lines from peripheral blood mononuclear cells for both patients and healthy donors. Three iPS cell lines were selected for each genotype. An increase in the copy number of genes does not always lead to an increasing in their expression level [8]. We showed that the *STAG2* expression was higher in all patient iPS cell lines compared to cells from healthy donors. Next, we differentiated the selected iPS cells into neural progenitor cells, and performed transcriptome analysis. The replicas converged well with each other according to PCA for all genes. We also discovered a number of transcriptional features in the patients' neural progenitor cells.

Conclusion: Xq25 microduplication leads to an increasing of the *STAG2* gene expression in iPS cells compared to cells with a normal karyotype. A number of transcriptional features were found in neural progenitor cells with *STAG2* gene duplication compared to cells with a normal karyotype.

Funding: The study was supported by the Ministry of Education and Science of Russian Federation, Agreement number No. 075-15-2024-539 dated April 24, 2024.

Список литературы/References

1. Meisenberg C., Pinder S.I., Hopkins S.R., Wooller S.K., Benstead-Hume G. et al. Repression of transcription at DNA breaks requires cohesin throughout interphase and prevents genome instability. *Mol Cell.* 2019;73(2):212-223.e7. doi 10.1016/j.molcel.2018.11.001
2. Merkenschlager M., Odom D.T. CTCF and cohesin: linking gene regulatory elements with their targets. *Cell.* 2013;152(6):1285-1297. doi 10.1016/j.cell.2013.02.029
3. Krantz I.D., McCallum J., DeScipio C., Kaur M., Gillis L.A. et al. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila* melanogaster Nipped-B. *Nat Genet* 2004;36(6):631-635. doi 10.1038/ng1364
4. Revenkova E., Focarelli M.L., Susani L., Paulis M., Bassi M.T. et al. Cornelia de Lange syndrome mutations in SMC1A or SMC3 affect binding to DNA. *Hum Mol Genet.* 2009;18(3):418-427. doi 10.1093/hmg/ddn369
5. Kumar R., Corbett M.A., Van Bon B.W.M., Gardner A., Woenig J.A., Jolly L.A. et al. Increased STAG2 dosage defines a novel cohesinopathy with intellectual disability and behavioral problems. *Hum Mol Genet.* 2015;24(25):7171-7181. doi 10.1093/hmg/ddv414
6. Leroy C., Jacquemont M.-L., Doray B., Lamblin D., Cormier-Daire V., Philippe A. et al. Xq25 duplication: the crucial role of the STAG2 gene in this novel human cohesinopathy. *Clin Genet.* 2016;89(1):68-73. doi 10.1111/cge.12567

7. Grigor'eva E.V., Kopytova A.E., Yarkova E.S., Pavlova S.V., Sorogina D.A., Malakhova A.A. et al. Biochemical characteristics of iPSC-N370S GBA variant carriers with and without parkinson's disease. *Int J Mol Sci*. 2023;24(5):4437. doi 10.3390/ijms24054437
8. Gridina M.M., Matveeva N.M., Fishman V.S., Menzorov A.G., Kizilova H.A., Beregovoy N.A. et al. Allele-specific biased expression of the CNTN6 gene in iPS cell-derived neurons from a patient with intellectual disability and 3p26.3 microduplication involving the CNTN6 gene. *Mol Neurobiol.* 2018;55(8):6533-6546. doi 10.1007/s12035-017-0851-5

doi 10.18699/bgrs2024-1.3-02

Необходимые параметры экструдеров и РНК-полимеразного комплекса для моделирования 3D структуры хроматина

Константинов В.В.[1, 2]*, Лагунов Т.А.[1, 2], Фишман В.С.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *v.konstantinov.work@gmail.com*

Ключевые слова: Hi-C; моделирование; РНК-полимераза; SMC

Мотивация и цель: В настоящий момент достаточно много известно о трехмерной структуре хроматина, начиная с влияния SMC-комплексов (конденсинов и когезинов) на доменную структуру и заканчивая активностью РНК-полимераз как основных факторов транскрипции. Однако вопрос об их взаимодействии раскрыт довольно слабо. Использование данных, собранных на модельном объекте, хромосомах типа ламповых щеток, позволило расширить наше понимание механизмов взаимодействия SMC-комплексов с Полимеразой II.

Методы и алгоритмы: Оценка физических параметров экструдеров и полимераз на основании литературных данных. Моделирование их взаимодействий на уровнях 1D и 3D с использованием программного пакета polychrom, анализ получаемых карт контактов. Сравнение модельных кривых P(s) с данными Hi-C эксперимента на одиночных ооцитах. Для получения зависимости вероятности контактов от расстояния использован программный пакет cooltools. Статистическая обработка результатов выполнена на языке python.

Результаты: В ходе изучения литературы были установлены допустимые границы параметров и ряд возможных механизмов взаимодействия. Так, был проверен механизм нетопологического перемещения конденсинов, что может объяснить способность конденсинов вытягивать хроматин несмотря на наличие препятствий большого размера на своем пути. Выполнено сравнение результатов моделирования для экструдеров с различной скоростью переключения направления вытягивания хроматина SMC-комплексом. Произведен подбор наиболее согласующихся с экспериментальными данными значений времени жизни активного экструдера и количества активных экструдеров. Исследована зависимость различных функций вероятности преодоления РНК-полимеразы и РНК-матрикса экструдером на формирование паттернов пространственных контактов, характерных для различной ориентации соседних генов.

Выводы: Проведенное исследование позволило воспроизвести ряд ранее наблюдаемых явлений в архитектуре хроматина за счет компьютерного моделирования. Это говорит в пользу топологического механизма перемещения SMC-комплексов и дает более полное представление об их взаимодействии с РНК-полимеразами.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда (проект № 22-14-00247).

Necessary parameters of extruders and RNA polymerase complex for modeling 3D chromatin structure

Konstantinov V.V.[1, 2]*, Lagunov T.A.[1, 2], Fishman V.S.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *v.konstantinov.work@gmail.com*

Key words: Hi-C, modelling; RNA-polymerase; SMC

Motivation and Aim: At the moment, quite a lot is known about the three-dimensional structure of chromatin, starting with the influence of SMC complexes (condensins and cohesins) on the domain structure and ending with the activity of RNA polymerases as the main transcription factors. However, the question of their interaction is rather poorly understood. The use of data collected on a model object, lampbrush chromosomes, has allowed us to expand our understanding of the mechanisms of interaction of SMC complexes with Polymerase II.

Methods and Algorithms: Estimation of physical parameters for extruders and polymerases based on literature data. Modeling of their interactions at the 1D and 3D levels using the polychrom software package, analysis of the resulting contact maps. Comparison of model P(s) curves with data from the Hi-C experiment on single oocytes. To obtain the dependence of the probability of contacts on distance, the cooltools software package was used. Statistical processing of the results was performed in Python.

Results: During the study of the literature, acceptable parameter limits and a number of possible interaction mechanisms were established. This is how the mechanism of non-morphological movement of condensins was tested, which may explain the ability of condensins to pull chromatin despite the presence of large obstacles in their path. The simulation results are compared for extruders with different speeds of switching the direction of chromatin extraction by the SMC complex. The selection of the values of the lifetime of the active extruder and the number of active extruders that are most consistent with the experimental data was carried out. The dependence of various probability functions of overcoming RNA polymerase and RNA matrix by an extruder on the formation of spatial contact patterns characteristic of different orientations of neighboring genes has been studied.

Conclusion: The study made it possible to reproduce a number of previously observed phenomena in chromatin architecture through computer modeling. This speaks in favor of the topological motive mechanism of SMC complexes and gives a more complete picture of their interaction with RNA polymerases.

Funding: The research was supported by RSF (project No. 22-14-00247).

doi 10.18699/bgrs2024-1.3-03

Поиск и описание хромосомных перестроек в хромосоме 1A у большой синицы (*Parus major*) с помощью метода Hi-C

Кораблёва С.Ю.[1, 2]*, Гридина М.М.[1], Торгашева А.А.[1], Нурисламов А.Р.[1], Задесенец К.С.[1], Малиновская Л.П.[1], Фишман В.С.[1]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *s.kozyreva@g.nsu.ru*

Ключевые слова: Hi-C; инверсия; геномные перестройки; большая синица

Мотивация и цель: Инверсии играют важную роль в эволюции хромосом, ограничивая поток генов в популяции. Они являются наиболее распространенными хромосомными перестройками в достаточно консервативных кариотипах птиц. Исследование инверсий важно для понимания механизмов видообразования и адаптации.

Ранее в популяции на территории Нидерландов был выявлен полиморфизм по комплексной перестройке в хромосоме 1A большой синицы (*Parus major*) [1]. Согласно данным анализа гетерозиготности и распределения количества однонуклеотидных полиморфизмов, перестройка предположительно включает в себя большую (более 90 % длины хромосомы) инверсию. Один из ее концов, вероятно, находится вблизи комплекса, для которого характерна вариация числа копий (copy number variants, CNV) [1]. При этом у 96 % гетерозигот по комплексной перестройке в хромосоме 1A число копий этого комплекса было существенно выше, чем у гомозигот по нормальной хромосоме 1A. Общая длина последовательностей, дополнительно присутствующих в перестроенной хромосоме, предположительно могла составлять около 3.5 Мб [1], что затрудняет определение точных границ инверсии.

С помощью методов цитогенетического анализа наши коллеги из Института цитологии и генетики СО РАН обнаружили полиморфизм, предположительно, по той же хромосомной перестройке в хромосоме 1A, у синиц из новосибирской популяции. Однако длина перестроенной хромосомы была примерно в 1.5 раза больше, чем нормальной, т. е. увеличение длины гомолога, несущего инвертированный аллель, составляло до 30 Мб – существенно больше, чем увеличение на 3.5 Мб, описанное раньше. Таким образом, структура этой перестройки – точные границы инверсии, а также генетическое содержание инсерции, неизвестны.

Методы и алгоритмы: Для исследования структуры сложной перестройки использовались методы Hi-C-секвенирования [2] и полногеномного секвенирования парными короткими прочтениями.

Для того чтобы определить нуклеотидную последовательность неизвестной инсерции, был разработан подход для поиска дополнительных последовательностей, отсутствующих в референсном геноме, с использованием данных полногеномного и Hi-C-секвенирования. В его основе лежит анализ отношения количества пар сегментов выравнивания, в которых одно из прочтений не картируется на геном, к парам, в которых оба прочтения выравниваются на геном.

Этот метод подходит для поиска неизвестных высокоповторенных последовательностей с помощью доступных методов секвенирования второго поколения, не прибегая к дорогостоящим методам третьего поколения.
Результаты: Впервые была построены Hi-C-карты для большой синицы дикого типа и с комплексной хромосомной перестройкой.
Нами были разработаны две версии алгоритма поиска неизвестных высокоповторенных последовательностей. Первый основан на сравнительном анализе Hi-C-данных животных с инверсией и дикого типа, второй – на анализе исключительно Hi-C-данных животных с инверсией и может применяться, если данные для животных дикого типа недоступны. С помощью этих алгоритмов была обнаружена последовательность длиной ~700 п. о., специфически амплифицированная у животных с инверсией. По предварительной оценке, суммарная длина копий данной последовательности составляет ~15 Мб. Эксперимент FISH с найденной последовательностью показал, что это повтор, диспергированный по перестроенному плечу хромосомы 1A.
С помощью секвенирования Nanopore удалось подтвердить результаты, полученные с помощью Hi-C секвенирования, и установить последовательность границ инверсии.
Выводы: с помощью метода Hi-C была построена карта контактов хроматина для синиц, гетерозиготных по комплексной перестройке в хромосоме 1A, и синиц дикого типа.
Уточнена сборка хромосомы 1A большой синицы, с нуклеотидной точностью определены границы инверсии и установлено, что границы инверсии фланкированы простыми повторами.
С помощью разработанного подхода и анализа копийности CNV обнаружены отсутствующие в референсном геноме синицы высокоповторенные последовательности ДНК, суммарная длина которых составляет около 3 Мб в геноме животных дикого типа и около 14 Мб в геноме животных с комплексной хромосомной перестройкой. Таким образом, объяснено увеличение гомолога хромосомы 1A с комплексной хромосомной перестройкой на 11 Мб по сравнению с гомологом дикого типа.
Финансирование: Работа выполнена при поддержке гранта РНФ 23-14-00182.

Dissecting the structure of the chromosomal rearrangements in chromosome 1A in great tits (*Parus major*) using Hi-C technique

Korableva S.Y.[1, 2]*, Gridina M.M.[1], Torgasheva A.A.[1], Nurislamov A.R.[1], Zadesenets K.S.[1], Malinovskaya L.P.[1], Fishman V.S.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *s.kozyreva@g.nsu.ru*

Key words: Hi-C; inversion; genome rearrangements; great tit

Motivation and Aim: Inversions play an important role in chromosome evolution, limiting gene flow within populations. They are the most common chromosomal

rearrangements in relatively conservative avian karyotypes. Studying inversions is crucial for understanding the mechanisms of speciation and adaptation.

Previously, polymorphism due to a complex rearrangement on chromosome 1A of the great tit (*Parus major*) population in the Netherlands [1] was identified. According to the analysis of heterozygosity and the distribution of single nucleotide polymorphisms, the rearrangement presumably involves a large inversion (>90 % of the chromosome length). One of the inversion breakpoint is likely to be near a complex characterized by copy number variation (CNV) [1]. Notably, in 96 % of individuals heterozygous for the complex rearrangement on chromosome 1A, the copy number of this complex was significantly higher than in individuals homozygous for the normal chromosome 1A. The total length of additional sequences present in the rearranged chromosome was estimated to be about 3.5 Mb [1], making it difficult to determine the exact boundaries of the inversion.

Using cytogenetic analysis methods, our colleagues from the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences discovered polymorphism, presumably due to the same chromosomal rearrangement on chromosome 1A, in a population of tits from Novosibirsk. However, the length of the rearranged chromosome was approximately 1.5 times longer than normal, indicating an increase in the length of the homolog carrying the inverted allele, up to 30 Mb – significantly more than the previously described increase of 3.5 Mb. Thus, the structure of this rearrangement – the exact inversion breakpoints as well as the genetic content of the insertion – remains unknown.

Methods and Algorithms: Hi-C sequencing methods [2] and whole-genome sequencing with paired short reads were employed to investigate the structure of the complex rearrangement. To determine the nucleotide sequence of the unknown insertion, an approach was developed for searching additional sequences absent in the reference genome, utilizing data from whole-genome and Hi-C sequencing. It is based on analyzing the ratio of read pairs, where one read is unmapped to the genome, to pairs where both reads are mapped. This method is suitable for identifying unknown highly repetitive sequences using available second-generation sequencing methods, without resorting to expensive third-generation methods.

Results: Hi-C maps were constructed for both wild-type great tits and those with complex chromosomal rearrangements for the first time. Two versions of an algorithm for identifying unknown highly repetitive sequences were developed. The first is based on comparative analysis of Hi-C data from animals with the inversion and wild type, while the second solely analyzes Hi-C data from animals with the inversion and can be applied when data for wild-type animals are unavailable. Using these algorithms, a sequence of approximately 700 base pairs long, specifically amplified in animals with the inversion, was discovered. The preliminary estimation of the copy number of this sequence is approximately 15 Mb. FISH experiment with the discovered sequence indicated it to be a repeat dispersed throughout the rearranged arm of chromosome 1A.

Sequencing with Nanopore technology confirmed the results obtained with Hi-C sequencing and established the sequence of inversion boundaries.

Conclusion: Using the Hi-C method, a chromatin contact map was constructed for tits heterozygous for the complex rearrangement on chromosome 1A and wild-type tits.

The assembly of chromosome 1A of the great tit has been refined, with nucleotide-level precision defining the boundaries of the inversion. It has been established that the inversion boundaries are flanked by simple repeats.

Using the developed approach and CNV copy number analysis, high-repeat DNA sequences absent in the reference genome of tits were discovered, with a total length of about 3 Mb in the genome of wild-type animals and about 14 Mb in the genome of animals with the complex chromosomal rearrangement on chromosome 1A. Thus, the increase in the homolog of chromosome 1A with a complex chromosomal rearrangement was explained by 11 Mb compared to the wild-type homolog.

Funding: The study is supported was supported by RSF grant No. 23-14-00182.

Список литературы/References

1. Da Silva V.H. The genomic complexity of a large inversion in great tits. *Genome Biol Evol*. 2019;11(7):1870-1881
2. Lieberman-Aiden E. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;32(5950):289-293. doi 10.1126/science.1181369

doi 10.18699/bgrs2024-1.3-04

Моделирование пространственной организации хромосом типа ламповых щеток

Лагунов Т.[1, 2]*, Константинов В.[1, 2], Гридина М.[1], Нурисламов А.[1, 2], Попов А.[1, 2], Куликова Т.[3], Красикова А.[3], Фишман В.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*

* *t.lagunov@g.nsu.ru*

Ключевые слова: Hi-C; биополимеры; хроматин; транскрипция

Мотивация и цель: Хромосомы типа ламповых щеток – это особая пространственная организации хромосом, образующаяся в ооцитах большинства животных и яйцекладущих млекопитающих (но не других млекопитающих). Для этой структуры характерны большой размер ядра, элонгированная форма хромосом и очень высокий уровень транскрипции. Выявление критически важных точек для формирования такой структуры с последующим моделированием пространственной организации представляет важную задачу для фундаментальной биологии.

Методы и алгоритмы: В основе исследования лежит анализ контактов хроматна одиночных ооцитов курицы, картированных при помощи технологии Hi-C, и моделирование взаимодействия экструдеров с РНК-полимеразой (базирующаяся на статье [1]). Первичный анализ Hi-C-данных проведен с использование пакета juicer-1.6. Зависимость вероятности контактов от расстояния получена инструментами модуля cooltools. RNA-seq-данные получены из статьи [2] и картированы при помощи bwa (ver. 0.7.17-r1188) с последующей разметкой генов пакетом stringtie (v2.2.2). Автоматическая разметка доменов хроматина выполнена инструментом lavaburst (v0.2.0). Моделирование пространственной организации выполнено при помощи polychrom (v0.1.0) [3]. Статистическая обработка производилась на языке python.

Результаты: На картах пространственных контактов хромосом типа ламповых щеток выделены доменные структуры. Показана высокая воспроизводимость расположения границ доменов между одиночными ооцитами курицы (по сравнению с одиночными ооцитами мыши). Большая часть границ доменов в ооцитах не содержит сайтов посадки белка CTCF, что исключает типичный для большинства куриных клеток механизм формирования доменов посредством взаимодействия когезина с белком CTCF. Большая часть границ доменов находится между генами, конвергентно ориентированными друг относительно друга. Конденсация хромосом при искусственном «выключении» транскрипции позволяет предположить активное взаимодействие SMC-комплексов с хроматином, а наличие активных генов в границах доменов приводит к предположению о формировании границ доменов хроматина посредством взаимодействия активных экструдеров с РНК-полимеразным комплексом. Произведена оценка параметров экструдеров, необходимых для адекватного воспроизведения паттернов пространственной организации хроматина куриных

ооцитов. Произведена оценка плотности посадки РНК-полимераз на основании данных RNA-seq. Проведено моделирование различных участков хроматина хромосом типа ламповых щеток. Показано соответствие наблюдаемых в моделях закономерностей с аналогичными закономерностями на карте пространственных контактов.
Выводы: Формирование хромосом типа ламповых щеток обусловлено, по большей части, высоким уровнем транскрипционной активности. Большое количество РНК-матрикса увеличивает персистентную длину участков активной транскрипции создавая петли, длина которых дополнительно увеличивается благодаря удалению нуклеосом. Большая часть паттернов контактов на Hi-C карте объясняется взаимодействием SMC-комплекса с активными полимеразами.
Финансирование: Работа выполнена при поддержке РНФ 22-14-00247.

Modeling spatial organization of lampbrush chromosomes

Lagunov T.[1, 2]*, Konstantinov V.[1, 2], Gridina M.[1], Nurislamov A.[1, 2], Popov A.[1, 2], Kulikova T.[3], Krasikova A.[3], Fishman V.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *St. Petersburg State University, St. Petersburg, Russia*

* *t.lagunov@g.nsu.ru*

Key words: Hi-C; biopolymers; chromatin; transcription

Motivation and Aim: Lampbrush chromosomes represent a distinctive spatial arrangement of chromatin formed in the oocytes of most animals and oviparous mammals (but not other mammals). This structure is characterized by the large nucleus size, elongated chromosome shapes, and a very high levels of transcription. The identification of essential mechanisms underlining formation of such a structure, followed by the modeling of spatial organization, is an important task for fundamental biology of chromatin.
Methods and Algorithms: The study is based on the analysis of single-cell Hi-C data produced from chicken oocytes and modeling of the interaction of the chromatin extruders with RNA polymerase (based on the article by Mirniy et al. 2023 [1]). Hi-C data was preprocessed using the juicer-1.6 package. The dependence of the contact probability from genomic distance was accessed using cooltools software. The RNA-seq data was obtained from [2], and alignment was performed using the bwa package (ver. 0.7.17-r1188), followed by transcriptome assembly using the stringtie package (v2.2.2). Automatic chromatin domain segmentation was performed by the lavaburst tool (v0.2.0). Spatial organization modeling was performed using polychrom (v0.1.0) [3]. Statistical processing was performed in python.
Results: We identified chromatin domains based on the analysis of the Hi-C maps. We show that reproducibility of domain boundaries between individual chicken oocytes is substantially than for single mouse oocytes. Most of the domain boundaries in chicken oocytes do not contain CTCF protein binding sites, which is inconsistent with the mechanism of domain formation through the interaction of cohesin with CTCF, typical for most chicken cell types. Most of the domain boundaries are located between

convergently transcribed genes. Based on the previous evidence of 1) lampbrush chromosome condensation caused by transcription inhibition and 2) the presence of SMC complexes on lampbrush chromosomes, we proposed formation of chromatin domain boundaries through the interaction of active extruders with the RNA polymerase complex. We determined parameters of extruders necessary to reproduce patterns detected in chicken oocyte Hi-C data. The separation between RNA polymerases of each transcription unit was estimated based on RNA-seq data. Using these praters, we developed models of several lampbrush chromosome loci and showed correspondence of the patterns observed in the models with the patterns on the corresponding Hi-C maps. *Conclusion*: The formation of lampbrush chromosomes occurs mainly due to very active transcription. A large amount of RNA matrix increases the persistent length of active transcription sites and results in the formation loops. The length of loops is further increased due to the nucleosomes removal. Most of the contact patterns on the Hi-C map were explained by the interaction of the SMC complex with active polymerases.
Funding: The study is supported by RSF 22-14-00247.

Список литературы/References

1. Banigan E.J., Tang W., Van Den Berg A.A. et al. Transcription shapes 3D chromatin organization by interacting with loop extrusion. *PNAS*. 2023;120(11):e2210480120
2. Krasikova A., Kulikova T., Schelkunov M., Makarova N., Fedotova A., Berngardt V., Maslova A., Fedorov A. The first chicken oocyte nucleus whole transcriptomic profile defines the spectrum of maternal mRNA and non-coding RNA genes transcribed by the lampbrush chromosomes. *bioRxiv*. 2024 doi 10.1101/2024.02.05.577752
3. Imakaev M., Goloborodko A. Hbbrandao mirnylab/polychrom: v0.1.0. mirnylab/polychrom. 2019. [https://zenodo.org/records/3579473]

doi 10.18699/bgrs2024-1.3-05

Трехмерная структура дальних поликомб-опосредованных контактов хроматина в нейронах коры головного мозга человека

Плетенев И.[1]*, Ваулин Н.[1], Молодова М.[1], Солдатенкова А.[1], Разин С.[2, 3], Ульянов С.[2, 3], Храмеева Е.[1]

[1] *Центр молекулярной и клеточной биологии, Сколтех, Москва, Россия*

[2] *Институт биологии гена РАН, Москва, Россия*

[3] *Московский государственный университет им. М.В. Ломоносова, биологический факультет, Москва, Россия*

* *ilya.pletenev@skoltech.ru*

Ключевые слова: хроматин; поликомб; нейроны; транскрипционные факторы; snm3C-seq

Мотивация и цель: Одной из особенностей трехмерной структуры хроматина нейронов коры головного мозга человека являются дальние контакты, предположительно формируемые белками группы поликомб [1]. Данные контакты были обнаружены в нейронах взрослого мозга, где они соединяют промоторы неактивных генов, кодирующих транскрипционные факторы, многие из которых участвуют в развитии нервной системы. Однако роль поликомбовых контактов в жизни клетки остается неясной. Прежде при изучении этих контактов мы использовали Hi-C карты нейронов, полученных из взрослого мозга и не разделенных на подтипы [1]. В данной работе мы использовали открытые данные о трехмерной структуре и транскриптоме одиночных нейронов [2–5], чтобы изучить возможные отличия паттерна дальних поликомб-опосредованных контактов в разных подтипах нейронов, а также на разных стадиях развития нейронов.

Методы и алгоритмы: В работе были проанализированы открытые данные о трехмерной структуре хроматина одиночных клеток, полученные методом snm3C-seq, а также открытые данные транскриптомики одиночных клеток snRNA-seq. Для анализа и визуализации данных были использованы библиотеки Python: cooler [6], cooltools [7], scanpy [8], а также программа HiGlass [9].

Результаты: Мы обнаружили, что паттерн поликомбовых контактов отличается в возбуждающих и тормозных нейронах. В частности, в тормозных нейронах, в отличие от возбуждающих, отсутствуют поликомбовые контакты с геном *SOX2*, кодирующим транскрипционный фактор, участвующий в развитии нервной системы, а сам ген имеет высокую экспрессию. Кроме того, интенсивность нейрональных поликомбовых контактов изменяется в ходе развития нервной системы, достигая максимума во взрослых нейронах.

Выводы: Полученные результаты позволяют предположить, что дальние поликомбовые контакты регулируют работу транскрипционных факторов в ходе развития мозга, а также могут быть важны для формирования разных подтипов нейронов.

Финансирование: Исследование поддержано грантом РНФ (№ 21-74-10102).

3D structure of long-range neuronal polycomb interactions in the human brain cortex

Pletenev I.[1]*, Vaulin N.[1], Molodova M.[1], Soldatenkova A.[1], Razin S.[2, 3], Ulianov S.[2, 3], Khrameeva E.[1]

[1] *Center for Molecular and Cellular Biology, Skoltech, Moscow, Russia*

[2] *Institute of Gene Biology, RAS, Moscow, Russia*

[3] *Lomonosov Moscow State University, Biological Faculty, Moscow, Russia*

** ilya.pletenev@skoltech.ru*

Key words: chromatin; polycomb; neurons; transcription factors; snm3C-seq

Motivation and Aim: Long-range polycomb interactions are the hallmark of neuronal chromatin 3D structure in the human brain cortex [1]. In mature neurons, these interactions bring together promoters of inactive genes that code for transcription factors, many of which are involved in nervous system development. However, the precise role of polycomb interactions in the neuronal development and function remains unclear. Previously, we used bulk Hi-C data to study neurons in the adult brain [1]. In this work, we used publicly available single-nucleus data on 3D structure and transcriptome of neurons [2–5] to study the pattern of polycomb interactions in the subtypes of mature neurons as well as in developing neurons.

Methods and Algorithms: In this work publicly available snm3C-seq and snRNA-seq data were analyzed. Python libraries: cooler [6], cooltools [7], and scanpy [8] – were used for analysis, and HiGlass software [9] – for visualization.

Results: We discovered that polycomb interactions occur at different loci in excitatory and inhibitory neurons. In particular, polycomb interactions with the *SOX2* gene, that code for neuronal transcription factor, are present in excitatory neurons, but not in inhibitory neurons. Accordingly, *SOX2* gene is expressed exclusively in inhibitory neurons, in line with repressive function of polycomb. We also found that the intensity of neuronal polycomb interactions increases in the course of brain development, reaching maximum in mature neurons.

Conclusion: Our results suggests that long-range polycomb interactions are involved in the regulation of neuronal transcription factors during brain development and may also be important for neuronal differentiation into subtypes.

Funding: The study is supported by the Russian Science Foundation grant (No. 21-74-10102).

Список литературы/References

1. Pletenev I.A., Bazarevich M., Zagirova D.R. et al. Extensive long-range polycomb interactions and weak compartmentalization are hallmarks of human neuronal 3D genome. *Nucleic Acids Res.* 2024;52(11):6234-6252. doi 10.1093/nar/gkae271
2. Tian W. et al. Single-cell DNA methylation and 3D genome architecture in the human brain. *Science.* 2023;382:eadf5357. doi 10.1126/science.adf5357
3. Siletti K. et al. Transcriptomic diversity of cell types across the adult human brain. *Science.* 2023;382:eadd7046. doi 10.1126/science.add7046
4. Heffel M. et al. Epigenomic and chromosomal architectural reconfiguration in developing human frontal cortex and hippocampus. *bioRxiv*. 2022. doi 10.1101/2022.10.07.511350
5. Herring C.A. et al. Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell*. 2022;185(23):4428-4447.e28. doi 10.1016/j.cell.2022.09.039

6. Abdennur N., Mirny L.A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics.* 2020;36(1):311-316. doi 10.1093/bioinformatics/btz540
7. Abdennur N. et al. Cooltools: Enabling high-resolution Hi-C analysis in Python. *PLoS Comp Biol.* 2020;20(5):e1012067. doi 10.1371/journal.pcbi.1012067
8. Wolf F., Angerer P., Theis F. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19(1):15. doi 10.1186/s13059-017-1382-0
9. Kerpedjiev P., Abdennur N., Lekschas F. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* 2018;19(1):125. doi 10.1186/s13059-018-1486-1

doi 10.18699/bgrs2024-1.3-06

Сборка генома обыкновенной ласки (*Mustela nivalis*) на уровне С-скаффолдов хромосомной длины

Тотиков А.[1, 2]*, Томаровский А.[1, 2], Перельман П.[1], Бульонкова Т.[3], Панов В.[4], Консорциум DNA Zoo[5], Графодатский А.[1], Кливер С.[6]

[1] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Институт систем информатики им. А.П. Ершова СО РАН, Новосибирск, Россия*

[4] *Институт систематики и экологии животных СО РАН, Новосибирск, Россия*

[5] *Центр архитектуры генома, Отделение молекулярной генетики и генетики человека, Медицинский колледж Бэйлора, Хьюстон, США*

[6] *Независимый исследователь, Копенгаген, Дания*

* *a.totickov1@gmail.com*

Ключевые слова: сборка генома; Hi-C; обыкновенная ласка; *Mustela nivalis*

Мотивация и цель: Обыкновенная ласка является широко распространенным мелким хищником, охватывающим весь Голарктический регион. Оценка генетического разнообразия различных популяций ласки вызывает большой интерес, поскольку отмечаются сокращения численности в некоторых локальных популяциях. Для более точного понимания генетического разнообразия необходимо наличие высококачественной сборки генома хромосомного уровня. Такие сборки возможно получить, используя данные о конформации хроматина (Hi-C). Используя данные Hi-C секвенирования, мы улучшили общедоступную сборку генома обыкновенной ласки на уровне отдельных скаффолдов до хромосомного уровня, а также сопоставили номенклатуру хромосомных скаффолдов с ранее полученными данными Zoo-FISH и G-бэндинга [1, 2].

Методы и алгоритмы: Общедоступная сборка генома *Mustela nivalis* (GCA_019141155.1) на уровне отдельных скаффолдов использовалась в качестве чернового варианта сборки [3]. Hi-C библиотеки были получены из 6 пассажа первичной линии клеток фибробластов ушного хряща особи, отловленной в Новосибирской области. Секвенирование Hi-C библиотек проводилось на базе консорциума “DNA Zoo”. Выравнивание Hi-C данных выполнено с использованием программы Juicer v1.6 [4]. Определение гаплотипических дупликаций проводилось программой purge_dups v1.2.6 [5]. Из данных были удалены те кандидатные регионы, в которых более 90 % последовательности было вовлечено в гаплотипическую дупликацию. Сборка генома осуществлялась с помощью программы 3d-dna v180419 [6]. Ручная коррекция проводилась с использованием Juicebox v2.16.00 [7]. Целостность сборки генома оценивалась с помощью программы BUSCO v5.4.2 с использованием базы данных OrthoDB v10 Mammalia [8, 9]. Полногеномные выравнивания выполнены программой LAST v1519 [10]. Визуализация данных выполнена с использованием программного пакета MAVR (https://github.com/mahajrod/MAVR).

Для анализа синтенных блоков использовались сборки геномов хромосомного уровня *Martes foina* (DNA Zoo), *N. vison* (GCF_020171115.1), *M. erminea* (GCF_009829155.1), *M. nivalis*, *M. lutreola* (GCF_030435805.1), *M. putorius furo*

(DNA Zoo) и *M. nigripes* (DNA Zoo). Множественное полногеномное выравнивание было выполнено с помощью программы Progressive Cactus v2.8.0 [11]. Извлечение синтенных блоков выполнено с помощью halSynteny v2.2 [12], с указанием параметров "--minBlockSize 50000 --maxAnchorDistance 50000". Полученные результаты были визуализированы с помощью программного пакета MACE (https://github.com/mahajrod/MACE).

Результаты: Сборка генома *M. nivalis* хромосомного уровня имеет размер 2.4 млрд п. н. и значение N50 138.3 млн п. н. Анализ целостности сборки генома с использованием базы данных Mammalia (всего 9226 ортологов) выявил 96 % цельных однокопийных ортологов и 3 % отсутствующих. Улучшенная сборка генома включает 21 хромосомный скаффолд (C-скаффолды), что соответствует гаплоидному набору хромосом в кариотипе *M. nivalis* (2*n*=42) [1, 2]. Наименование C-скаффолдов в сборке *M. nivalis* проведено с использованием ранее предложенной номенклатуры хромосом [1, 2]. Для этого были сопоставлены: полногеномные выравнивания сборки *M. nivalis* на сборку *Neogale vison* (GCF_020171115.1, 2*n*=30), сборку *M. lutreola* (GCF_030435805.1, 2*n*=38) и сборку *M. erminea* (GCF_009829155.1, 2*n*=44, рис. 1B), а также результаты Zoo-FISH с использованием зондов на основе хромосом *N. vison* на хромосомы *M. lutreola* и *M. nivalis* [1], и G-banded кариотипы этих видов [2]. После идентификации C-скаффолды были отнесены к соответствующим номерам хромосом с использованием префикса "chr" (рис. 1A).

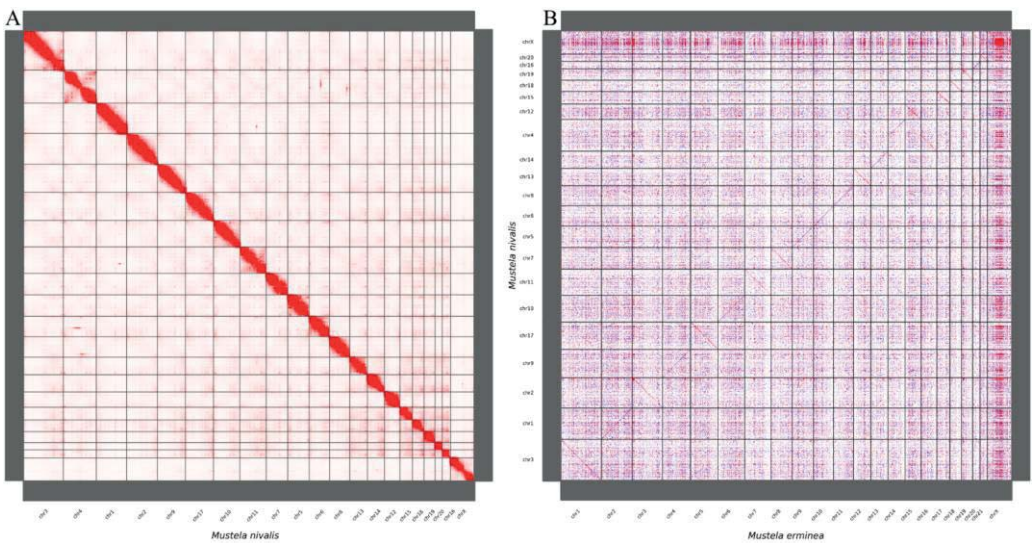


Рис. 1. (A) Карта плотности Hi-C контактов сборки генома обыкновенной ласки с 21 хромосомным скаффолдом. (B) Полногеномное выравнивание хромосомной сборки обыкновенной ласки на хромосомную сборку генома горностая (*M. erminea*)

При анализе синтенных блоков (рис. 2) обнаружены ранее цитогенетически выявленные транслокации [1]. Например, робертсоновская транслокация двух акроцентрических хромосом, наблюдаемая у большинства видов рода *Mustela*, слияние которых образует хромосому 1 у *M. lutreola* и *M. nigripes*. Также выявлено несколько крупных инверсий размером более 1 млн п. н., которые ранее подтверждены не были.

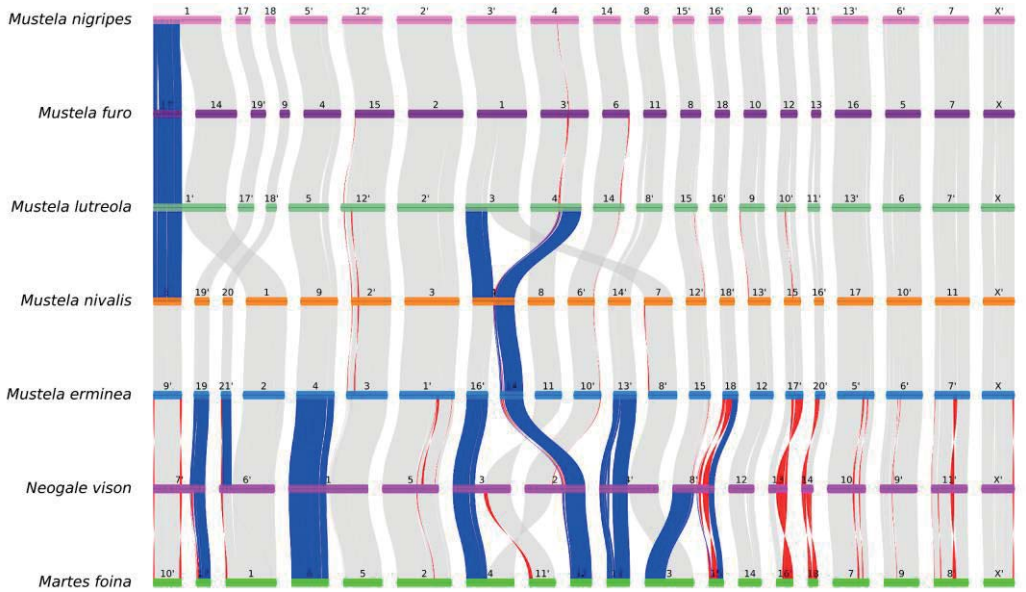


Рис. 2. Синтенные блоки между видами подсемейства Mustelinae и *Martes foina*. Инверсии и транслокации размером не менее 1 млн п.н. выделены красным и синим цветами соответственно

Выводы: Используя Hi-C данные, мы улучшили сборку генома *M. nivalis* до хромосомного уровня. Были выявлены многочисленные инверсии размером более 1 млн п.н., которые ранее определены не были. Полученная сборка генома может использоваться в различных популяционно-генетических исследованиях, позволяя получать более точные результаты.

Финансирование: За возможность проведения данной работы мы благодарим: Olga Dudchenko, Ruqayya Khan и Erez Aiden (Консорциум “DNA Zoo”, США); Karol Zub (MRI, Польша); Inês Miranda, Liliana Farelo и Jose Melo-Ferreira (CIBIO, Португалия); Юлию Бутакову (ИМКБ СО РАН, Россия). Исследование поддержано грантом РНФ (№ 22-24-01076).

Chromosome-level genome assembly of the Least Weasel (*Mustela nivalis*)

Totikov A.[1, 2]*, Tomarovsky A.[1, 2], Perelman P.[1], Bulyonkova T.[3], Panov V.[4], DNA Zoo Consortium[5], Graphodatsky A.[1], Kliver S.[6]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *A.P. Ershov Institute of Informatics Systems, SB RAS, Novosibirsk, Russia*

[4] *Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia*

[5] *The Center for Genome Architecture, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA*

[6] *Independent scientist, Copenhagen, Denmark*

* *a.totickov1@gmail.com*

Key words: genome assembly; Hi-C; least weasel; *Mustela nivalis*

Motivation and Aim: The Least Weasel is a widely distributed small predator encompassing the entire Holarctic region. The assessment of genetic diversity among different populations of Least Weasels is of great interest due to observed declines in some local populations. To achieve a more accurate understanding of genetic diversity, high-quality genome assemblies at the chromosome level are necessary. Such assemblies can be obtained using chromatin conformation data (Hi-C). Utilizing Hi-C sequencing data, we have improved the publicly available genome assembly of the *M. nivalis* to the chromosome level and have also aligned the nomenclature of chromosomal scaffolds with previously obtained Zoo-FISH and G-banding data [1, 2].
Methods and Algorithms: The publicly available genome assembly of *M. nivalis* (GCA_019141155.1) at the scaffold level was utilized as a draft assembly [3]. Hi-C libraries were generated from the 6th passage of the primary fibroblast cell line derived from ear cartilage of an male individual captured in the Novosibirsk region. Hi-C library sequencing was conducted by the "DNA Zoo" consortium. Alignment of Hi-C data was performed using Juicer v1.6 software [4]. Haplotypic duplications were identified using purge_dups v1.2.6 software [5]. Candidate regions where more than 90 % of the sequence was involved in haplotypic duplication were removed. Genome assembly was carried out using 3d-dna v180419 software [6]. Manual curation was conducted using Juicebox v2.16.00 software [7]. Genome assembly completeness was assessed using BUSCO v5.4.2 software with the OrthoDB v10 Mammalia database [8, 9]. Whole-genome alignments were performed using LAST v1519 software [10]. Data visualization was conducted using the MAVR software package (https://github.com/mahajrod/MAVR). The synteny blocks analysis utilized chromosome-level genome assemblies at the *Martes foina* (DNA Zoo), *N. vison* (GCF_020171115.1), *M. erminea* (GCF_009829155.1), *M. nivalis*, *M. lutreola* (GCF_030435805.1), *M. putorius furo* (DNA Zoo), and *M. nigripes* (DNA Zoo). Multiple whole-genome alignments were conducted using Progressive Cactus v2.8.0 [11]. Syntenic blocks were extracted using halSynteny v2.2 [12] with parameters set to "--minBlockSize 50000 --maxAnchorDistance 50000". The resulting outcomes were visualized using the MACE software package (https://github.com/mahajrod/MACE).
Results: The improved genome assembly has a size of 2.4 Gbp and an N50 value of 138.3 Mbp. BUSCO analysis for the *M. nivalis* assembly, we identified 96 % complete and single-copy BUSCOs from the Mammalia database (9,226 BUSCOs), with only 3 % missing. The improved genome assembly consists of 21 chromosomal scaffolds (C-scaffolds), corresponding to the haploid chromosome set in the karyotype of *M. nivalis* (2*n*=42) [1, 2]. The nomenclature of C-scaffolds in the *M. nivalis* assembly was established using previously proposed chromosome nomenclature [1, 2]. This involved comparing whole-genome alignments of the *M. nivalis* assembly to the assembly of *Neogale vison* (GCF_020171115.1, 2*n*=30), the assembly of *M. lutreola* (GCF_030435805.1, 2*n*=38), and the assembly of *M. erminea* (GCF_009829155.1, 2*n*=44), as well as the results of Zoo-FISH probes based on *N. vison* chromosomes on *M. lutreola* and *M. nivalis* chromosomes [1], and G-banded karyotypes of these species [2]. Following the identification of C-scaffolds, they were assigned to corresponding chromosome numbers using the prefix "chr" (Fig. 1).

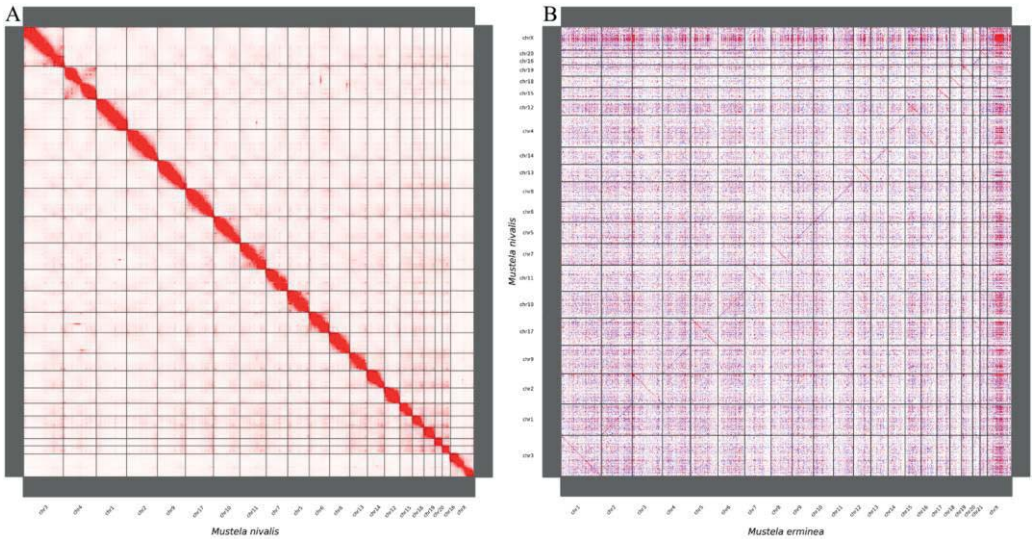


Fig. 1. (A) Hi-C contact density map of the Least Weasel genome assembly showing 21 chromosome length scaffolds. (B) Whole genome alignment of the Least Weasel chromosome length assembly to the ermine's (*M. erminea*) assembly

During the analysis of synteny blocks (Fig. 2), previously cytogenetically identified translocations were detected [1]. For instance, the Robertsonian translocation involves two acrocentric chromosomes, observed in most species of the genus *Mustela*, the fusion of which forms chromosome 1 in *M. lutreola* and *M. nigripes*. Additionally, several large inversions exceeding 1 Mbp were identified.

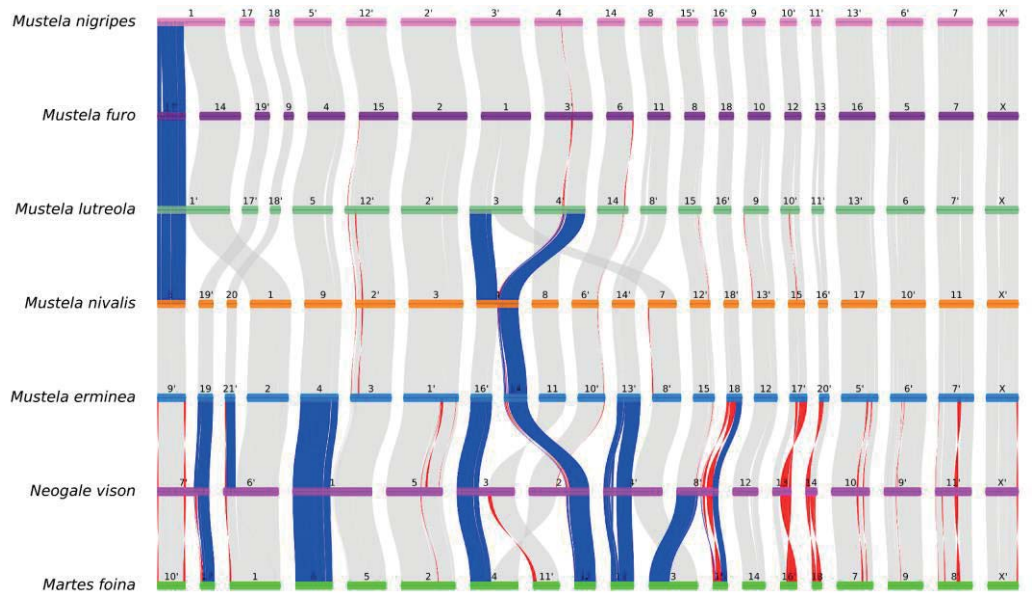


Fig. 2. Synteny blocks between species of the Mustelinae subfamily and *Martes foina*. Inversions and translocations of 1 Mbp and larger are highlighted in red and blue, respectively

Conclusion: Using Hi-C data, we have successfully improved the genome assembly of *M. nivalis* to the chromosome level. We detected numerous inversions larger than 1 Mbp that were previously undetected. The resulting genome assembly can be utilized in various population genetics studies, enabling more precise results to be obtained.

Funding: We acknowledge for making this genome assembly possible: Olga Dudchenko, Ruqayya Khan and Erez Aiden (Consortium "DNA Zoo", USA); Karol Zub (MRI, Poland); Inês Miranda, Liliana Farelo, Jose Melo-Ferreira (CIBIO, Portugal); and Yulia Butakova (IMCB SB RAS, Russia). The study is supported by the RSF grant (No. 22-24-01076).

Список литературы/References

1. Graphodatsky A.S., Yang F., Perelman P.L. et al. Comparative molecular cytogenetic studies in the order Carnivora: Mapping chromosomal rearrangements onto the phylogenetic tree. *Cytogenet Genome Res*. 2002;96(1-4):137-145. doi 10.1159/000063032
2. Graphodatsky A., Perelman P., O'Brien S.J. Atlas of Mammalian Chromosomes. John Wiley & Sons, Incorporated, 2020. doi 10.1002/9781119418061
3. Miranda I., Giska I., Farelo L. et al. Museomics dissects the genetic basis for adaptive seasonal colouration in the least weasel. *Mol Biol Evol*. 2021;38(10):4388-4402. doi 10.1093/molbev/msab177
4. Durand N.C., Shamim M.S., Machol I. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 2016;3(1):95-98. doi 10.1016/j.cels.2016.07.002
5. Guan D., McCarthy S.A., Wood J., Howe K., Wang Y., Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 2020;36(9):2896-2898. doi 10.1093/bioinformatics/btaa025
6. Dudchenko O., Batra S.S., Omer A.D. et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science.* 2017;356(6333):92-95. doi 10.1126/science.aal3327
7. Durand N.C., Robinson J.T., Shamim M.S. et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst*. 2016;3(1):99-101. doi 10.1016/j.cels.2015.07.012
8. Manni M., Berkeley M.R., Seppey M. et al. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38(10):4647-4654. doi 10.1093/molbev/msab199
9. Kriventseva E.V., Kuznetsov D., Tegenfeldt F. et al. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res*. 2019;47(D1):D807-D811. doi 10.1093/nar/gky1053
10. Kiełbasa S.M., Wan R., Sato K., Horton P., Frith M.C. Adaptive seeds tame genomic sequence comparison. *Genome Res*. 2011;21(3):487-493. doi 10.1101/gr.113985.110
11. Armstrong J., Hickey G., Diekhans M. et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature.* 2020;587(7833):246-251. doi 10.1038/s41586-020-2871-y
12. Krasheninnikova K., Diekhans M., Armstrong J. et al. halSynteny: a fast, easy-to-use conserved synteny block construction method for multiple whole-genome alignments. *GigaScience*. 2020;9(6):giaa047. doi 10.1093/gigascience/giaa047

doi 10.18699/bgrs2024-1.3-07

Нарушение межнуклеосомных взаимодействий приводит к заметным изменениям в архитектуре генома

Ульянов С.В.[1], Храмеева Е.Е.[2], Михалева Е.А.[3], Разин С.В.[1], Шевелев Ю.Я.[3]*

[1] *Институт биологии гена РАН, Москва, Россия*

[2] *Сколковский институт науки и технологий, Сколково, Россия*

[3] *Национальный исследовательский центр «Курчатовский институт», Москва, Россия*

* *shevelev@img.ras.ru*

Ключевые слова: ацетилирование гистонов; acidic parch; архитектура генома; LANA

Мотивация и цель: В неацетилированном хроматине взаимодействие N-конца гистона H4 одной нуклеосомы с участком acidic patch, образованным димером H2A/H2B другой нуклеосомы, приводит к агрегации нуклеосом и компактизации хроматиновой нити [1, 2]. В то же время ацетилирование N-конца гистона H4 предотвращает его связывание с участком acidic patch, нарушая тем самым межнуклеосомные взаимодействия и деконденсируя хроматин [3–6]. С участком acidic patch могут также связываться многочисленные белки хроматина (например, вирусный антиген LANA, нуклеопорин Elys и многие другие) [7, 8]. Поскольку эти белки конкурируют с неацетилированным N-концом гистона H4 за связывание с тем же интерфейсом, они могут мешать образованию межнуклеосомных взаимодействий. Целью данного исследования было выявление изменений в архитектуре генома, происходящих в ответ на искусственно вызванное нарушение межнуклеосомных взаимодействий.

Методы и алгоритмы: Hi-C проводили с использованием фермента Hind III, как было описано ранее [9].

Результаты: Чтобы исследовать, как нарушение взаимодействий между нуклеосомами влияет на архитектуру генома, мы применили два различных подхода. В первом из них мы транзиентно трансфицировали клетки S2 дрозофилы конструкцией, экспрессирующей пептид LANA, который способен связываться с участком acidic patch, тем самым предотвращая взаимодействие между этим интерфейсом и положительно заряженным N-концом неацетилированного гистона H4 из других нуклеосом. Мы ожидали, что экспрессия LANA приведет к нарушению агрегации неацетилированных нуклеосом в топологически ассоциированных доменах (ТАДах), состоящих из неактивного хроматина. Во втором подходе мы транзиентно трансфицировали клетки S2 конструкцией, экспрессирующей мутантную форму гистона H2A, лишенную аминокислот, образующих участок acidic patch, что также должно предотвращать агрегацию неацетилированных нуклеосомных нитей.

Мы провели Hi-C на клетках, экспрессирующих эти белковые факторы, с последующим секвенированием Hi-C библиотек. Идентифицированные ТАДы были разделены на четыре группы (A, B, C, D) в зависимости от соотношения в них активных и неактивных типов хроматина, где ТАДы группы A содержали наибольшую долю активного хроматина, а ТАДы группы D – наименьшую его долю. Затем мы подсчитали среднюю частоту контактов (ACF) внутри каждого ТАДа. На рис. 1 показаны изменения значений ACF в клетках, экспрессирующих

LANA или мутантный гистон H2A (H2Amut), по сравнению с контрольными клетками. При экспрессии LANA или H2Amut значения ACF и, следовательно, плотность упаковки ТАДов группы D, состоящих из неактивного хроматина, снижается. Таким образом, ингибирование связывания неацетилированного N-конца гистона H4 одной нуклеосомы с acidic patch интерфейсом другой нуклеосомы приводит к разрыхлению общей пространственной организации неактивных ТАДов. В то же время экспрессия обоих белковых факторов приводит к увеличению плотности упаковки ТАДов группы A, состоящих из активного высокоацетилированного хроматина (рис. 1).

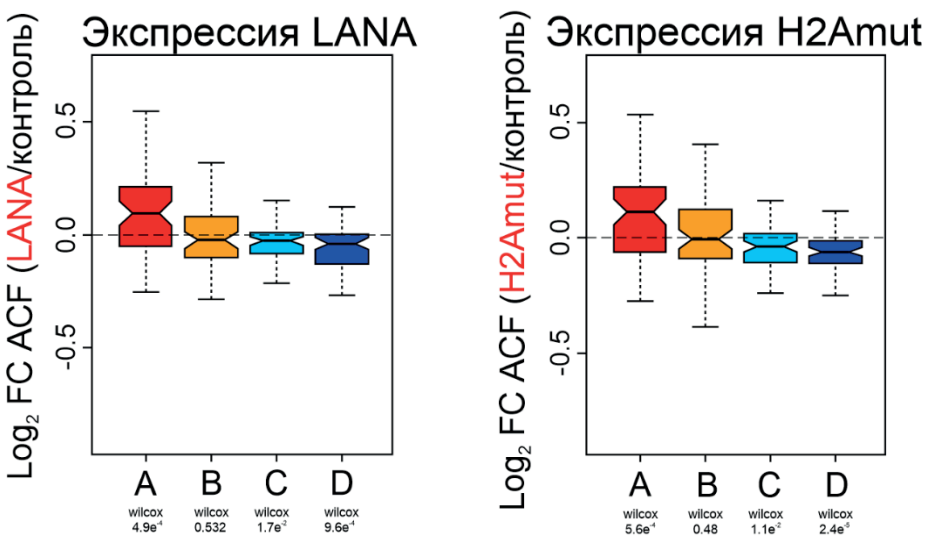


Рис. 1. Экспрессия LANA или мутантного гистона H2A в клетках S2 приводит к снижению плотности упаковки неактивных ТАДов и увеличению плотности упаковки активных ТАДов. Log_2(fold change ACF) в ТАДах четырех групп после экспрессии в клетках S2 LANA (левая панель) или мутантного гистона H2A, не способного образовывать acidic patch (правая панель). *P*-values определяли с помощью теста Вилкоксона

Было показано, что с участком acidic patch ацетилированных нуклеосом может связываться нуклеопорин Elys [8, 10], который, в свою очередь, может привлекать комплекс ремоделирования хроматина PBAP [11], деконденсирующий хроматин. В таком случае блокирование участка acidic patch с помощью LANA или отсутствие участка acidic patch в гистоне H2Amut может предотвращать связывание комплекса Elys/PBAP, тем самым делая хроматин более конденсированным.

Выводы: Наша работа показала, что как межнуклеосомные взаимодействия, так и связывание определенных белковых комплексов с нуклеосомами являются ключевыми факторами, определяющими трехмерную организацию хроматина в интерфазном ядре. Нарушение любого типа взаимодействий приводит к заметным изменениям в архитектуре генома.

Финансирование: Исследование выполнено в рамках государственного задания НИЦ «Курчатовский институт».

Disruption of inter-nucleosomal interactions leads to notable changes in genome architecture

Ulianov S.V.[1], Khrameeva E.E.[2], Mikhaleva E.A.[3], Razin S.V.[1], Shevelyov Y.Y.[3]*

[1] *Institute of Gene Biology, RAS, Moscow, Russia*

[2] *Skolkovo Institute of Science and Technology, Skolkovo, Russia*

[3] *National Recearch Centre "Kurchatov Institute", Moscow, Russia*

* *shevelev@img.ras.ru*

Key words: histone acetylation; acidic patch; genome architecture; LANA

Motivation and Aim: In the non-acetylated chromatin, the interaction of histone H4 tail of one nucleosome with the acidic patch interface formed by H2A/H2B dimer of another nucleosome leads to the aggregation of nucleosomes and to the compactization of chromatin fiber [1, 2]. At the same time, the acetylation of H4 tail prevents its binding with the acidic patch, thus disrupting inter-nucleosomal interactions and decondensing chromatin [3–6]. Numerous chromatin proteins (for example, viral latency-associated nuclear antigen (LANA), nucleoporin Elys, and many others) may also bind with the acidic patch [7, 8]. Since these proteins compete with the non-acetylated histone H4 tail for binding with the same interface, they may interfere with inter-nucleosomal interactions. The goal of this study was to reveal changes in genome architecture occurring in response to artificially induced disturbance of inter-nucleosomal interactions.

Methods and Algorithms: High-throughput chromosome conformation capture (Hi-C) was performed using Hind III digestion enzyme as described previously [9].

Results: To examine how the disturbance of interactions between nucleosomes affects genome architecture, we applied two different approaches. In the first one, we transiently transfected *Drosophila* S2 cells with the construct expressing LANA peptide, which is capable of binding to the acidic patch region, thereby preventing interactions between this interface and the positively charged tail of non-acetylated histone H4 from other nucleosomes. We expected that LANA expression would lead to the disruption of aggregation of non-acetylated nucleosomes in topologically associating domains (TADs) composed from the inactive chromatin. In the second approach, we transiently transfected S2 cells with the construct expressing mutant form of histone H2A, lacking amino acids forming the acidic patch region, which should also prevent the aggregation of non-acetylated nucleosomal arrays.

We performed Hi-C in cells expressing these protein factors, followed by sequencing of Hi-C libraries. The identified TADs were divided into four groups (A, B, C, D), depending on the proportion of active and inactive chromatin types within them, where group A TADs contained the largest proportion of active chromatin, while group D TADs – the lowest proportion of it. Next, average contact frequency (ACF) for each TAD was calculated. Figure 1 shows changes in ACF values in cells expressing LANA or mutant histone H2A (H2Amut) relative to control cells. Upon the expression of LANA or H2Amut, ACF values and, consequently, the packaging density of group D TADs, consisting of inactive chromatin, decreases. Thus, inhibition of binding of non-acetylated histone H4 terminus from one nucleosome with the acidic patch interface of another nucleosome leads to the loosening of general spatial organization of inactive

TADs. At the same time, the expression of both protein factors leads to the increase in packaging density of group A TADs, consisting of active, highly acetylated chromatin (Fig. 1).

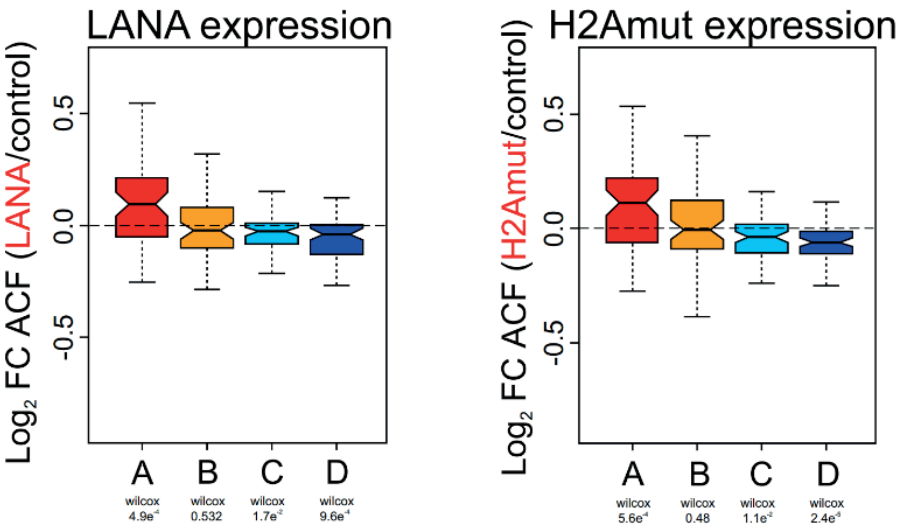


Fig. 1. Overexpression of LANA or mutant histone H2A in S2 cells leads to a decrease in the packaging density of inactive TADs and to an increase in the packaging density of active TADs. Log_2(fold change ACF) in TADs of four groups after overexpression in S2 cells of LANA (left panel) or mutant histone H2A, unable to form acidic patch interface (right panel). *P*-values were determined in the Wilcoxon test

It was shown that the acidic patch interface of acetylated nucleosomes may be bound by nucleoporin Elys [8, 10], which, in turn, may recruit the chromatin remodeling complex PBAP [11], known to decondense chromatin. Then, blocking of acidic patch by LANA, or the lack of acidic patch in the H2Amut can prevent binding of Elys/PBAP complex, thus making chromatin more condensed.

Conclusion: Our work has shown that both inter-nucleosomal interactions and binding of certain protein complexes to nucleosomes are the key factors determining the three-dimensional organization of chromatin in interphase nucleus. Disruption of any type of interactions leads to notable changes in genome architecture.

Funding: The study was carried out within the framework of the state task of NRC "Kurchatov Institute".

Список литературы/References

1. Luger K., Mäder A.W., Richmond R.K., Sargent D.F., Richmond T.J. Crystal structure of the nucleosome core particle at 2.8. A resolution. *Nature.* 1997;389:251-260. doi 10.1038/38444
2. Sinha D., Shogren-Knaak M.A. Role of direct interactions between the histone H4 Tail and the H2A core in long range nucleosome contacts. *J Biol Chem.* 2010;285:16572-16581. doi 10.1074/jbc.M109.091298
3. Shogren-Knaak M., Ishii H., Sun J.M., Pazin M.J., Davie J.R., Peterson C.L. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science.* 2006;311:844-847. doi 10.1126/science.1124000

4. Allahverdi A., Yang R., Korolev N., Fan Y., Davey C.A., Liu C.F., Nordenskiöld L. The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic Acids Res.* 2011;39:1680-1691. doi 10.1093/nar/gkq900
5. Zhang R., Erler J., Langowski J. Histone acetylation regulates chromatin accessibility: role of H4K16 in inter-nucleosome interaction. *Biophys J.* 2017;112:450-459. doi 10.1016/j.bpj.2016.11.015
6. Otterstrom J., Castells-Garcia A., Vicario C., Gomez-Garcia P.A., Cosma M.P., Lakadamyali M. Super-resolution microscopy reveals how histone tail acetylation affects DNA compaction within nucleosomes in vivo. *Nucleic Acids Res.* 2019;47:8470-8484. doi 10.1093/nar/gkz593
7. Barbera A.J., Chodaparambil J.V., Kelley-Clarke B., Joukov V., Walter J.C., Luger K., Kaye K.M. The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science.* 2006;311:856-861. doi 10.1126/science.1120541
8. Skrajna A., Goldfarb D., Kedziora K.M., Cousins E.M., Grant G.D., Spangler C.J., Barbour E.H., Yan X., Hathaway N.A. et al. Comprehensive nucleosome interactome screen establishes fundamental principles of nucleosome binding. *Nucleic Acids Res.* 2020;48:9415-9432. doi 10.1093/nar/gkaa544
9. Ulianov S.V., Khrameeva E.E., Gavrilov A.A. et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* 2016;26:70-84. doi 10.1101/gr.196006.115
10. Kobayashi W., Takizawa Y., Aihara M., Negishi L., Ishii H., Kurumizaka H. Structural and biochemical analyses of the nuclear pore complex component ELYS identify residues responsible for nucleosome binding. *Commun Biol.* 2019;2:163. doi 10.1038/s42003-019-0385-7
11. Kuhn T.M., Pascual-Garcia P., Gozalo A., Little S.C., Capelson M. Chromatin targeting of nuclear pore proteins induces chromatin decondensation. *J Cell Biol.* 2019;218:2945-2961. doi 10.1083/jcb.201807139

doi 10.18699/bgrs2024-1.3-08

Изменение 3D организации хроматина в локусе приводит к нарушению экспрессии гена SIX3 в раннем нейральном развитии

Хабарова А.[1]*, Гридина М.[1, 2], Степанчук Я.[1, 2], Нурисламов А.[1, 2], Кораблева С.[1, 2], Крепиши А.[3], Коста С.[3], Бертола Д.[3], Оливейра Д.[3], Карвало Л.[3], Фишман В.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Университет Сан-Паулу, Сан-Паулу, Бразилия*

** khabarova@bionet.nsc.ru*

Ключевые слова: 3D геном; хромосомные перестройки; задержка развития; SIX3

Мотивация и цель: Нашим коллегами для консультации и медико-генетического исследования был направлен пациент (девочка 16 лет) с лицевыми диморфизмами, задержкой развития, аутистическими чертами и тугоухостью. Схожие клинические проявления обнаружены у ее брата и отца. В результате генетического исследования была обнаружена делеция 95 КБ 2p21 (chr2:45187369-45282639, GRCh37), затрагивающая только ген SIX2. Известно, что члены семейства транскрипционных факторов SIX участвуют в эмбриональном морфогенезе черепа и почечной системы. Ген SIX2 широко экспрессируется у плода в конце первого триместра, в том числе в конечностях, среднем ухе и почках, но его экспрессия происходит в ограниченном диапазоне тканей у взрослых. Известно, что мутации SIX2 у человека приводят к синдрому лобно-носовой дисплазии [1] и порокам развития косточек верхнего века и среднего уха, что приводит к потере слуха [2], наблюдаемой у пациента. На данный момент не показано участие SIX2 в развитии головного мозга, что могло бы объяснить описанную задержку ментального развития. Однако подобные нарушения описаны для другого гена – SIX3, расположенного в том же локусе, что и SIX2, на расстоянии 14 т.п.о. от границ делеции. SIX3 – фактор транскрипции, который участвует в развития глаз, носа, женской репродуктивной системы и переднего мозга. Различные нарушения гена SIX3 приводят к тяжелому пороку развития головного мозга – голопрозэнцефалии [3]. Локус SIX2/SIX3 с точки зрения трехмерной организации хроматина состоит из двух доменов, причем гены SIX2 и SIX3 лежат в разных доменах. Обнаруженная у пациента делеция затрагивает не только ген SIX2, но и границу, разделяющую эти домены. Мы предполагаем, что делеция, обнаруженная у пациентов, может приводить к изменению пространственной организации окружающего участка ДНК, что в свою очередь может вызвать нарушение регуляции гена SIX3. Дополнительные исследования, включая анализ пространственной организации данного локуса в клетках пациентов и анализ экспрессии генов SIX2 и SIX3, могут дать дополнительную поддержку этой гипотезе.

Методы и алгоритмы: Поскольку экспрессия гена SIX3 необходима на самых ранних этапах развития нервной системы для правильного формирования переднего мозга [4], в качестве целевой популяции клеток, на которой мы будем

оценивать эффекты хромосомной перестройки данного локуса, мы используем клетки нейральных предшественников. На первом этапе мы планируем получить индуцированные плюрипотентные стволовые клетки из периферических мононуклеаров крови пациентов, а затем дифференцировать их в клетки нейральных предшественников по протоколу Muratore et al. [5]. В клетках пациентов и здоровых контролей мы планируем оценить экспрессию гена SIX3.
Результаты: Мы получили и проанализировали данные РНК-секвенирования в трех образцах с делецией гена SIX2 и девяти образцах нормы. Была показана достоверная корреляция между экспрессией генов SIX2 и SIX3. Так, во всех образцах с делецией SIX2 экспрессия SIX3 значительно повышалась.
Выводы: В клетках нейральных предшественников была показана корреляция между экспрессией гена SIX2 и SIX3: делеция гена SIX2 приводит к значительному повышению экспрессии гена SIX3.
Финансирование: Исследование поддержано грантом РНФ 22-14-00247.

Disruption in the 3D organization of chromatin at the locus leads to impaired expression of the SIX3 gene in early neural development

Khabarova A.[1]*, Gridina M.[1, 2], Stepanchuk Y.[1, 2], Nurislamov A.[1, 2], Korableva S.[1, 2], Krepishi A.[3], Costa S.[3], Bertola D.[3], Oliveira D.[3], Carvalho L.[3], Fishman V.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *University of San Paulu, San Paulu, Brazil*

** khabarova@bionet.nsc.ru*

Key words: 3D genome; chromosomal rearrangements; developmental delay; SIX3

Motivation and Aim: Our colleagues for consultation and medical genetic research referred a patient (a 16-year-old girl) with facial dimorphisms, developmental delay, autistic traits and hearing loss. Her brother and father had similar clinical manifestations. A genetic study revealed a deletion of 95 KB 2p21 (chr2:45187369-45282639, GRCh37) affecting only the SIX2 gene. It is known that members of the SIX family of transcription factors are involved in the embryonic morphogenesis of the skull and renal system. The SIX2 gene is widely expressed in the fetus at the end of the first trimester, including in the extremities, middle ear and kidneys, but its expression occurs in a limited range of tissues in adults. Currently, SIX2 mutations in humans have been shown to lead to frontal nasal dysplasia syndrome [1] and malformations of the bones of the upper eyelid and middle ear, which leads to hearing loss [2] observed in the patient. The participation of SIX2 in the development of the brain has not been shown and gene deletion could not explain the described delay in mental development. However, similar disorders have been described for another gene, SIX3, located at the same locus as SIX2 at a distance of 14 bp from the boundaries of the deletion. SIX3 is a transcription factor that is involved in the development of the eyes, nose, female reproductive system and forebrain. Various disorders of the SIX3 gene lead to severe brain malformation – holoprosencephaly [3]. From the point of view of the three-dimensional organization of chromatin, the SIX2/SIX3 locus consists of two domains, with the SIX2 and SIX3 genes

lying in different domains. The deletion detected in the patient affects not only the SIX2 gene, but also the boundary separating these domains. We suggest that the deletion found in patients may lead to a change in the spatial organization of the surrounding DNA region, which in turn may cause a violation of the regulation of the SIX3 gene. Additional studies, including analysis of the spatial organization of this locus in patients' cells and analysis of the expression of genes SIX2 and SIX3, may provide additional support for this hypothesis.

Methods and Algorithms: Since the expression of the SIX3 gene is necessary at the earliest stages of the development of the nervous system for the proper formation of the forebrain [4], we use neural progenitor cells as the target cell population on which we will evaluate the effects of chromosomal rearrangement of this locus. At the first stage, we plan to obtain induced pluripotent stem cells from peripheral mononuclear blood of patients, and then differentiate them into neural progenitor cells according to the protocol Muratore et al. [5]. In the cells of patients and healthy controls, we plan to evaluate the expression of the SIX3 gene.

Results: We obtained and analyzed RNA sequencing data in three SIX2 gene deletion samples and 9 normal samples. A significant correlation was shown between the expression of the SIX2 and SIX3 genes: in all samples with the SIX2 deletion, the expression of SIX3 was significantly increased.

Conclusion: A correlation between the expression of the SIX2 and SIX3 gene was shown in neural progenitor cells: deletion of the SIX2 gene leads to a significant increase in the expression of the SIX3 gene.

Funding: The Russian Science Foundation supported the study (No. 22-14-00247).

Список литературы/References

1. Hufnagel R.B., Zimmerman S.L. et al. A new frontonasal dysplasia syndrome associated with deletion of the SIX2 gene. *Am J Med Genet A*. 2016;170A(2):487-491. doi 10.1002/ajmg.a.37441
2. Guan J., Wang D., Cao W. et al. SIX2 haploinsufficiency causes conductive hearing loss with ptosis in humans. *J Hum Genet*. 2016;61(11):917-922. doi 10.1038/jhg.2016.86
3. Laflamme C., Filion C., Labelle Y. Functional characterization of SIX3 homeodomain mutations in holoprosencephaly: interaction with the nuclear receptor NR4A3/NOR1. *Hum Mutat*. 2004;24(6):502-508. doi 10.1002/humu.20102
4. Gestri G., Carl M., Appolloni I. et al. Six3 functions in anterior neural plate specification by promoting cell proliferation and inhibiting Bmp4 expression. *Development*. 2005;132(10):2401-2413. doi 10.1242/dev.01814
5. Muratore C.R., Srikanth P. et al. Comparison and optimization of hiPSC forebrain cortical differentiation protocols. *PLoS One*. 2014 28;9(8):e105807. doi 10.1371/journal.pone.0105807

doi 10.18699/bgrs2024-1.3-09

The role of cohesion in providing the physical properties of the nucleus and DNA breaks repair

Battulin N.[1]*,Yunusova A.[1], Kabirova E.[1, 2], Ryzhkova A.[1], Khabarova A.[1], Maltseva E.[2], Smirnov A.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *battulin@bionet.nsc.ru*

Key words: chromatin, cohesin, protein degradation, mechanical properties of the nucleus

Motivation and Aim: The cohesin protein complex ensures the packaging of chromatin in the nuclear space in the form of loops—topologically associated domains. It is known that this level of packaging is involved in the implementation of genome functions—gene transcription, establishing and maintaining the interaction of the enhancer with the target gene and isolating the regulatory contexts of different genes. At the same time, cohesin-mediated chromatin looping contributes to genome stability because it can facilitate the repair of double-strand DNA breaks by restraining DNA strands after the break. By structuring chromatin, cohesin contributes to the mechanical properties of the nucleus. However, it remains unknown how significant the contribution of cohesin is to this process, and in particular in response to mechanical stress.

Methods and Algorithms: The solution to this issue is associated with a limitation: cohesin is necessary for normal cell division, that is, obtaining a cell line with deletion of the cohesin complex proteins is difficult. This problem can be overcome by a protein degradation system triggered in response to the plant hormone auxin. With this approach, we are able to disrupt the assembly of the cohesin complex, causing degradation of the RAD21 subunit just one hour after auxin addition. This model allows us to study the rapid effect of the absence of cohesin. We carried out work on HCT116 tumor cells with RAD21 degron.

Results: In this work, we ask the question of the contribution of cohesin to cell resistance to mechanical stress. As a first step in studying this issue, we assessed the ability of the cell nucleus in the absence of cohesin to resist mechanical deformations that occur during normal cell activity. To do this, we used the scratch wound assay: we disrupted the integrity of the cell monolayer ("wound") and, using time-lapse microscopy, observed the healing of the "wound" by cells within 24 hours. Comparison of efficiency and rate of wound healing showed no differences between groups of cohesin-depleted cells and control HCT116 cells.

Conclusion: Thus, we can conclude that cohesin depletion does not affect the migratory properties of cells when moving along a flat surface. Interestingly, according to the literature, when cohesin is depleted, predominantly normal gene expression occurs.

It can be assumed that under normal conditions, the loss of cohesin does not lead to serious disruptions in the functioning of the cell until the moment of division. In continuation of our work, our model will allow us to study how the absence of cohesin affects the resistance of cells to mechanical stress and DNA repair.

Funding: The work was carried out with financial support from the Russian Science Foundation (project No. 23-74-00055).

doi 10.18699/bgrs2024-1.3-10

Fundamental and applied 3D genomics

Fishman V.[1, 2]

[1] *Institute of Cytology and Genetics, Novosibirsk, Russia*

[2] *Artificial Intelligence Research Institute, AIRI, Moscow, Russia*

minja-f@yandex.com

Key words: 3D genomics; genome regulation; chromatin architecture; structural variants

This year marks the 15th year of the pioneering publication that introduced a genome-wide chromosome conformation capture technique, setting a milestone in 3D genomics. Over the past decade and a half, these techniques have become essential tools for addressing a wide array of genomic challenges. Our session on "Fundamental and Applied 3D Genomics" will explore key questions in this dynamic field. Initially, we will explore recent discoveries concerning the mechanisms of chromatin organization. Studies of our group, utilizing genetically modified mouse models, have elucidated a complex relationship between gene expression regulation and the maintenance of local chromatin domains, emphasizing the role of CTCF in this intricate process. Contrarily, we will also present cases from specific cell types demonstrating that CTCF binding does not always correlate with chromatin domain formation. Additionally, we introduce new physical models that illuminate how active transcription influences the formation of these CTCF-independent structures. Our findings significantly enhance our understanding of 3D genomics and its implications for regulatory mechanisms underlying human monogenic diseases and cancer development.

Another key focus of our research involves the application of chromatin conformation technologies to detect structural variants and improve genome assembly. We will highlight innovative computational methods and experimental workflows that facilitate the detection of both structural and single-nucleotide variants in the human genome. These methods have proven crucial in cancer research and the molecular diagnosis of monogenic diseases. Furthermore, the advent of single-cell techniques has enabled the application of these technologies to low-input human samples, ultimately making them suitable for analyzing human embryo biopsies. These advancements allow for the effective identification of balanced chromosomal translocations in clinical settings.

Funding: The study is supported by RSF Grant 22-14-00247.

doi 10.18699/bgrs2024-1.3-11

Cohesin degradation effect on mechanical properties of human colon tumor cells

Kabirova E.[1, 2]*, Yunusova A.[1], Khabarova A.[1], Battulin N.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *e.kabirova@g.nsu.ru*

Key words: 3D genome; chromatin; auxin-inducible degradation; mechanical stability; cell migration

Motivation and Aim: Cohesin mediates chromatin loop extrusion forming topologically associated domains which establish promoter-enhancer interactions. That way cohesin contributes to fulfill one of the primary genome functions – gene expression. In addition, cohesin-mediated chromatin extrusion can facilitate DNA repair by restraining DNA strands after the break, thus contributing to genome stability. Since cohesin influences the genome through its spatial structure, it is of interest whether cohesin provides mechanical stability to the nucleus. We aimed to assess cohesin influence on mechanical properties of a cell, in particular cell resistance to mechanical stress.

Methods and Algorithms: Obtaining a cell line lacking cohesin is essential to determine the cohesin role in the process (to compare the effects with and without cohesin). However, it is associated with a limitation: cohesin is necessary for a normal cell division, thus creating a cohesin knockout is difficult. The issue can be overcome by a protein degradation system induced by the plant hormone auxin. Utilizing the approach, we were able to disrupt the assembly of the cohesin complex, caused by degradation of the RAD21 subunit (RAD21 degron). This model allows us to study the rapid effect of the cohesin absence, just one hour after auxin introduction to the system. We carried out work on a human colon tumor cell line (HCT116) with RAD21 degron [1] and HCT116 without degron as a control. To determine the ability of a cell nucleus in an absence of cohesin to resist mechanical deformations, we performed (a) "scratch wound" assay for analyzing cell migration on flat surface, (b) transwell assay for analyzing active cell migration in a serum gradient.

Results: Comparison of wound healing by cohesin-depleted and control HCT116 revealed no differences in efficiency and rate between the groups. For transwell, HCT116 cells migration through 8 um pores was more efficient than through 5 um pores (4 hours of migration through 8 um resembled 24 hours of migration through 5 um). But the difference of migration efficiency between cohesin-depleted and control HCT116 was mild.

Conclusion: Thus, we can conclude that cohesin depletion does not affect the migratory properties of cells when moving along a flat surface or migrating in a gradient of serum. Though there is a possibility that cohesin influence on mechanical properties of a nucleus is low relative to other features such as cytoskeleton or nuclear lamina. To further study the nucleus mechanical properties we are planning to distinguish contribution of each of the elements (cohesin, cytoskeleton, lamina).

Funding: The Russian Science Foundation supported the study (No. 23-74-00055).

References

1. Yesbolatova A., Saito Y., Kitamoto N., Makino-Itou H. et al. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat Commun.* 2020;11(1):5701. doi 10.1038/s41467-020-19532-z

doi 10.18699/bgrs2024-1.3-12

Monogenic and multigenic transcription loops in lampbrush chromosomes of birds

Krasikova A.*, Kulikova T., Maslova A., Berngardt V., Plotnikov V.
Saint-Petersburg State University, St. Petersburg, Russia
** alla.krasikova@gmail.com*

Key words: chromomere-loop complex; hypertranscription; lampbrush chromosomes; nascent RNA; nuclear RNA-seq; oocyte nucleus; RNA-FISH; transcription loops; transcription unit

Motivation and Aim: Genome expression can be regulated by the speed of transcription. The rate of nascent RNA synthesis is unusually high in lampbrush chromosomes. This is most likely achieved by the tight coverage of transcribed genes by elongating RNA polymerases [1]. Recently, we established that at the lampbrush chromosome stage, avian oocytes accumulate enormous amounts of mRNAs, long non-coding RNAs and small non-coding RNAs that are synthesised in the oocyte nucleus [2]. We concluded that in avian oocytes, hypertranscription on the lateral loops of giant lampbrush chromosomes is a primary mechanism to synthesise large amounts of maternal RNAs for each of the transcribed genes. Since we detected unspliced or partially spliced nuclear transcripts for many genes, we aimed to detect nascent transcripts and characterize the organision of transcription units *in situ*.
Methods and Algorithms: Here, we used an RNA-FISH protocol that allows us to map nascent gene transcripts on the lateral loops of lampbrush chromosomes [3]. FISH probes were prepared from the chicken BAC-clone library CHORI-261 (https://bacpacresources.org/chicken261.htm) to the selected genomic regions. BAC-clone based probes were labelled by nick-translation as described in [4]. Lampbrush chromosomes were isolated microsurgically from the nuclei of chicken diplotene oocytes. FISH on lampbrush chromosome preparations was performed according to DNA/RNA and DNA/DNA+RNA hybridisation procedures. Control experiments were carried out to assess the quality of spreads and the integrity of nascent transcripts on lampbrush chromosome preparations [3]. RNA-seq data for chicken lampbrush stage cytoplasmic and nuclear RNA fractions are available at NCBI Sequence Read Archive (SRR23800660-SRR23800671) [2].
Results: By RNA-FISH we visualised transcription units of a number of protein-coding genes on chicken lampbrush chromosomes. Specifically, we described transcription loops for RBFOX1, RBMS3, CCSER1, GRID2, INPP5A and PARD3B genes, some of them being involved in embryo development and maternal RNA stability. For all studied examples, we observed complete correspondence of transcript detection by nuclear RNA-seq and by RNA-FISH with nascent transcripts. Genomic regions that are non-transcribed according to RNA-seq data were found within chromomeres. Such genomic regions often lack annotated genes (so-called gene deserts) and may contain untranscribed genes. Interestingly, a lateral loop inserted into a chromomere may contain several individual transcription units separated by untranscribed chromatin knots.
Using RNA-FISH with probes at the beginning, middle and end of the transcribed gene, we demonstrated that introns are removed from lampbrush transcripts before the end of the transcription unit. Visualisation of co-transcriptional splicing in case of RBFOX1

and RBMS3 genes by light microscopy is accompanied by detection of correctly spliced maternal mRNAs deposited in the oocyte cytoplasm. We also conclude that the length of the transcription loops depends on the length of the gene, particularly the length of introns. In the late oocyte nucleus, transcriptional output tends to decrease while the boundaries of transcription units are maintained. Consistently, RNA-FISH data revealed nascent gene transcripts on the drastically shortened lateral loops in late stage oocytes.
Conclusion: We argue that nuclear RNA-seq profile allows predicting the chromomere-loop pattern of lampbrush chromatin domains. Genomic regions containing a number of actively transcribed genes would form a transcription loop, while genomic regions without transcription units would be packed into chromomeres or chromatin knots. We observed both monogenic and multigenic transcription loops organised as follows: (1) a single transcribed gene on a relatively long lateral loop, (2) a long gene with lengthy introns on an extremely long lateral loop with multiple RNP matrix gradients, (3) many genes of different polarity transcribed in a single lateral loop and separated by chromatin knots. We conclude that lampbrush chromosomes provide an especially useful system for studying thousands of transcription loops as well as the packaging of transcriptionally active chromatin and nascent RNAs. Since hypertranscription is not unique to the oocyte nucleus, lampbrush chromosomes are a valuable model for studying the mechanisms of global transcriptome upregulation.
Funding: The research was supported by the Russian Science Foundation (grant #19-74-20075) and was performed using the equipment of the Genomics Core Facility (Skoltech) and Resource Center “Molecular and Cell Technologies” (Saint-Petersburg State University).

References

1. Krasikova A., Fishman V., Kulikova T. Lampbrush chromosome studies in the post-genomic era. *BioEssays*. 2023;45:2200250. doi 10.1002/bies.202200250
2. Krasikova A., Kulikova T., Schelkunov M., Makarova N., Fedotova A., Berngardt V., Maslova A., Fedorov A. The first chicken oocyte nucleus whole transcriptomic profile defines the spectrum of maternal mRNA and non-coding RNA genes transcribed by the lampbrush chromosomes. *bioRxiv*. 2024. doi 10.1101/2024.02.05.577752
3. Kulikova T., Krasikova A. RNA-FISH on lampbrush chromosomes: visualization of individual transcription units. In: Cytogenetics and molecular cytogenetics. Boca Raton: CRC Press, 2022;307-318
4. Krasikova A., Kulikova T., Rodriguez Ramos J.S., Maslova A. Assignment of the somatic A/B compartments to chromatin domains in giant transcriptionally active lampbrush chromosomes. *Epigenet Chromatin*. 2023;16:24. doi 10.1186/s13072-023-00499-2

doi 10.18699/bgrs2024-1.3-13

CTCF and the evolution of genome spatial organization

Nurislamov A.[1, 2]*, Popov A.[1, 2], Gridina M.[2], Shadskiy A.[1, 2], Fishman V.[1, 2]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *a.nurislamov@g.nsu.ru*

Key words: chromatin; 3D genomics; epigenetics; evolution

Motivation and Aim: The CCCTC-binding factor (CTCF) is a crucial insulator protein in bilaterian animals. While the structure of CTCF is conserved, its functions can vary across different taxa. In vertebrates, CTCF is involved in forming chromatin loops via a cohesin/CTCF-dependent extrusion mechanism, which is a primary factor in the formation of topologically associated domains (TADs) in these animals. This mechanism has not been observed in other bilaterian groups despite the conserved structure of the protein. Therefore, it is of great interest to identify factors that determine the presence or absence of cohesin/CTCF-dependent loop extrusion in bilaterian animals.

Methods and Algorithms: We conducted an evolutionary analysis of CTCF amino acid sequences, focusing particularly on the N-terminal domain. This domain exhibits the most variable structure among bilaterian animals and is critically important for chromatin loop formation in vertebrates. To compare the emergence time in evolution of crucial amino acid motifs for cohesin interaction, we employed multiple protein alignment techniques. By aligning CTCF sequences from a wide range of bilaterian taxa, we were able to identify conserved and divergent motifs and determine their evolutionary timelines.

Results: Our analysis identified N-terminal domain sequences specific to jawed vertebrates, which are absent in jawless vertebrates and cephalochordates. We discovered that the KTYQR motif emerged specifically in jawed vertebrates, indicating a potential evolutionary adaptation linked to their chromatin organization mechanisms. Additionally, we found that the YXF motif is present in CTCF across all studied bilaterian groups, except for *Trichinella spiralis*, where CTCF likely loses its function. In vertebrates, the YXF motif in CTCF is essential for chromatin loop formation as it facilitates the interaction between CTCF and the loop extrusion complex. This widespread presence of the YXF motif across bilaterian taxa underscores its fundamental role in CTCF functionality, whereas the emergence of the KTYQR motif in jawed vertebrates suggests additional evolutionary adaptations enhancing CTCF's role in chromatin organization in these species.

Conclusion: Our findings suggest that the ability of vertebrate CTCF to participate in chromatin loop formation is influenced not only by the presence of the YXF motif but also by other factors. This highlights the complexity of the mechanisms underlying chromatin organization and the evolutionary divergence of CTCF functions in bilaterian animals.

Funding: The study is supported by RSF grant No. 22-14-00247.

doi 10.18699/bgrs2024-1.3-14

The influence of SINEs as potential regulatory element on chromatin structures in human brain cancer cells

Panchenko D.D.
Applied Genomics Laboratory, SCAMT Institute, ITMO University, St. Petersburg, Russia
panchenko@scamt-itmo.ru

Key words: SINEs; RNAseq; HiC analysis; chromatin reorganization; network approach; network centralities

Motivation and Aim: High-grade gliomas are notably aggressive and malignant brain tumors [1, 2]. Although they account for about one-fourth of all brain tumors, they demonstrate a high tendency for recurrence, often leading to a short median progression-free survival of only 1.8 months after recurrence [3, 4]. Despite surgical resection being a primary treatment strategy, the invasive nature of these tumors into surrounding healthy brain tissue makes complete removal challenging, thereby reducing the likelihood of successful long-term management. Moreover, despite significant advancements in medical research, high-grade gliomas continue to pose a formidable challenge, with many aspects of their underlying genetic and epigenetic mechanisms remaining elusive [5]. One area of study is the role of SINEs in genomic regulation [6–8]. This group of repeats include two large families: Alu and MIR which perform different functions in the genome and have a potential effect on transcription processes. This study aims to explore how SINEs affect the spatial organization of the genome and their subsequent impact on gene regulation in gliomas. Employing a comprehensive approach that includes Hi-C analysis to assess the 3D genome architecture, transcriptomic analysis to identify gene dysregulation, and genomic analyses to map SINE distribution, this research seeks to identify patterns and mechanisms by which SINE elements contribute to the development and progression of cancer.
Methods and Algorithms: The methodology employed in this study follows a systematic framework comprising several analytical stages to address the research objectives (Fig. 1). Initially, transcriptomic analysis is conducted to identify genes implicated in cancerogenesis, utilizing approaches such as differential gene expression analysis to discern patterns of gene dysregulation. Subsequently, the spread of SINEs within the genome is studied, with a focus on identifying SINE elements located in proximity to gene regions. This involves genomic analyses to characterize the distribution and abundance of SINEs across the human genome. Hi-C analysis investigates differences in genome organization between samples, aiming to identify correlations with the findings from transcriptomic and SINE analysis. Also in this study was performed a new method of investigation interaction between distinct chromosomal regions with high resolution. Closeness and betweenness centralities were used for detecting potential regulatory bins which could take part in chromatin reorganization. Finally, visualization to highlight key regions of the genome and explore their potential 3D structure, offering insights into the functional implications of SINE-mediated genome organization.
Results: Genomic analysis has revealed that 8.43 % of the entire genome is occupied by SINEs, with 26.2 % located within gene regions and 31.1 % in promoter regions extending from –50Kb to the transcription start site (TSS). Detailed investigations were

focused on SINEs within these promoter regions. Transcriptomic analysis of two diffuse intrinsic pontine glioma (DIPG) samples, compared to a control sample from healthy brain tissue, revealed differential expression in 297 genes. These genes are primarily involved in pathways responsible for the regulation of DNA-templated transcription and nervous system development, showing upregulation in DIPG. In contrast, when comparing glioblastoma multiforme (GBM) samples with controls, pathways associated with cell division were predominantly affected. Additionally, a downregulation of the Keratin Filament pathway (GO:0045095) was observed in both DIPG and GBM samples. Comparative Hi-C analysis across these samples yielded an average correlation coefficient of 0.39 between DIPG patients and the control, and 0.43 between GBM samples and the control. RNA-seq analysis corroborated these findings, indicating a closer genetic expression between the control and GBM than with DIPG samples. In the final stage of analysis, chromatin regions were examined for 5000 bp bins showing high contact frequency with adjacent bins and containing SINEs. For instance, a bin at 52,165,000–52,170,000 within the Intermediate Filament (IntFil) gene cluster on chromosome 12 was identified. This bin showed extensive contacts with neighboring bins in the 52–54 Mb region and contained a previously identified SINE with a high identity score. Additional regions of interest with highly conserved SINEs were identified on chromosomes 1, 5, 11, 17, and X, emphasizing the widespread impact of these elements across the genome.

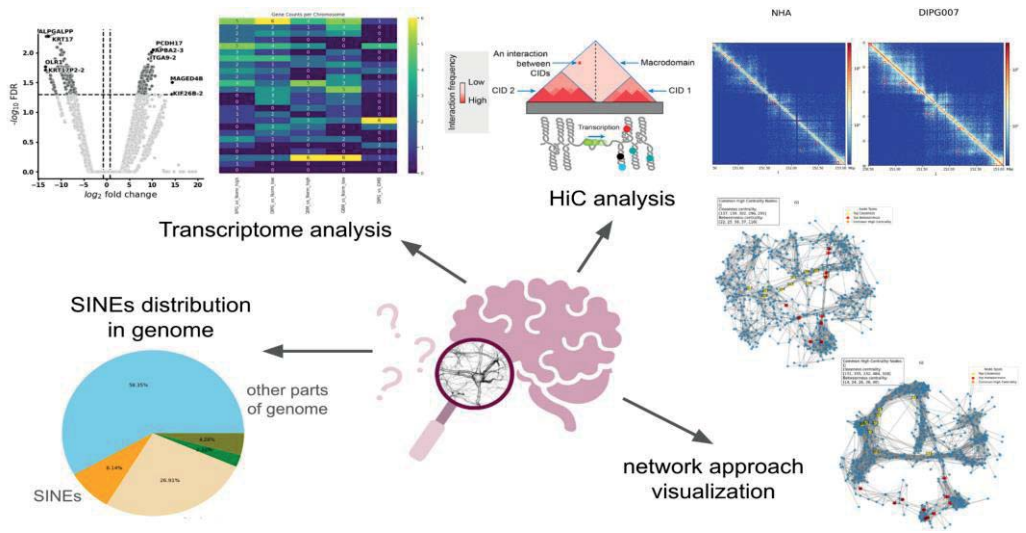


Fig. 1. Stages of research. Step1. Identification of differential gene expression between healthy and tumor cell samples. Step2. SINEs distribution in the genome. Step3. 3D chromatin organization. Step4. Visualization of the most significant chromatin regions

Conclusion: This study elucidates the role of SINEs in influencing the genomic architecture and regulatory landscapes in high-grade gliomas. By integrating Hi-C, transcriptomic, and genomic analyses, the study provides novel insights into how SINEs may drive the complex pathophysiology of these aggressive tumors. Ultimately, these findings could pave the way for developing targeted therapeutic strategies that address the genetic and epigenetic underpinnings of glioma recurrence and progression.

References

1. Louis D.N. et al. The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro Oncol*. 2021;23(8):1231-1251. doi 10.1093/neuonc/noab106
2. Xu Y. et al. Efficacy and safety of pharmacotherapy for recurrent high-grade glioma: a systematic review and network meta-analysis. *Front Pharmacol*. 2023;14:1191480. doi 10.3389/fphar.2023.1191480
3. McKinnon C., Nandhabalan M., Murray S.A., Plaha P. Glioblastoma: Clinical presentation, diagnosis, and management. *BMJ*. 2021;374:n1560. doi 10.1136/bmj.n1560
4. Wang J. et al. Epigenomic landscape and 3D genome structure in pediatric high-grade glioma. *Sci Adv*. 2023;7(39):eabg4126. doi 10.1126/sciadv.abg4126
5. Liao X. et al. Repetitive DNA sequence detection and its role in the human genome. *Commun Biol*. 2023;6(1):954. doi 10.1038/s42003-023-05322-y
6. Zhang X.-O., Pratt H., Weng Z. Investigating the potential roles of SINEs in the human genome. *Annu Rev Genomics Hum Genet*. 2021;22:199-218. doi 10.1146/annurev-genom-111620-100736
7. Ramsoomair C.K. et al. The epitranscriptome of high-grade gliomas: a promising therapeutic target with implications from the tumor microenvironment to endogenous retroviruses. *J Transl Med*. 2023;21(1):893. doi 10.1186/s12967-023-04725-z
8. Douville C. et al. Seq-ing the SINEs of central nervous system tumors in cerebrospinal fluid. *Cell Rep Med*. 2023;4(8):101148. doi 10.1016/j.xcrm.2023.101148

doi 10.18699/bgrs2024-1.3-15

The tissue-specific outcomes of TAD boundary disruption in Unc5b-Slc29a3 locus

Salnikov P.[1, 2]*, Stepanchuk Y.[1, 2], Yan A.[1, 2], Torgunakov N.[2], Belokopytova P.[1, 2], Fishman V.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Artificial Intelligence Research Institute (AIRI), Moscow, Russia*

* *paul.salnikov@gmail.com*

Key words: CTCF binding sites; cis-regulating elements; TAD boundary; *Slc29a3*; *Unc5b*

Motivation and Aim: In vertebrates, interphase chromatin is organized into structures known as Topologically Associated Domains (TADs) and smaller loop structures. These structures are highly conserved across species, and disturbances in TAD structures can drastically affect gene expression, potentially leading to disease. Researchers often focus on developmental genes that exhibit binary types of expression regulation, which may render them relatively insensitive to most TAD perturbation effects. Conversely, the impact of TAD disturbances on housekeeping genes, which are active in all cell types, is often overlooked. We propose that merging regulatory regions containing these ubiquitously expressed genes could lead to competition among promoters for enhancer activity, potentially altering gene expression levels.

In this study, we examine genes with broad activity profiles that are separated into two distinct TADs by a robust insulatory boundary. Specifically, we explore the consequences of CTCF binding sites deletion at *Slc29a3*/*Unc5b* locus, which includes genes crucial for development and viability, organized into two separate TADs by a conserved insulatory boundary. We hypothesize that disruptions at the TAD boundary could lead to misregulation of these genes, providing insight into how genome organization affects transcriptional activity across different tissues.

Methods and Algorithms: To assess gene expression alterations resulting from CTCF binding site deletions, we employed a robust experimental design involving F1 of mice with specific CTCF-bs deletions with wild-type CAST mouse strains. This allowed us to differentiate between allelic transcripts due to distinct exonic SNPs. Thus, the effects of the CTCF binding site deletions could be specifically analyzed within the same tissue samples by comparing the expression of CAST and C57BL/6 alleles. This design significantly reduces experimental biases as both alleles are subjected to the regulation except spatial cis-environment. Then we generated cDNA using UMI-barcoded, transcript-specific primers. Subsequent next-generation sequencing (NGS) analysis of these products enabled the discrimination and quantification of each allele. Using this strategy, we profiled allelic expression levels of hybrid groups carrying wild-type or deletion allele along with CAST allele, assessing differences between their levels of expression.

Results: For our experiments, we used previously obtained mouse line with deletions of CTCF binding sites (CTCF-bs) at TAD boundary region in *Slc29a3*/*Unc5b* locus. Hi-C maps derived from animal samples revealed noticeable changes at the subTAD level. We confirmed that loops within the *Slc29a3* TAD, originally anchored by the now-

deleted CTCF-bs, were disrupted in analyzed tissues (liver, kidney, and cerebellum). Typically, in wild-type animals, these loops facilitate connections from the TAD boundary to the CTCF-bs within the Cdh23 gene. Despite these disruptions, the rest of the inner loop structure orchestrating this TAD remained intact.

Significantly, the chromatin region extending from the TAD boundary to the Cdh23 promoter, which encompasses the *Slc29a3* gene, developed increased spatial interactions with the entire Unc5b TAD. This indicates that this DNA region changed its association to an adjacent TAD, with the TAD boundary shifting towards the Cdh23 promoter.

Furthermore, we observed the emergence of new, aberrant long-range interactions. These newly formed loops extend from the inner CTCF-binding sites at the Cdh23 gene to the outer borders of the *Unc5b* TAD, an area harboring silent non-coding RNA genes. Remarkably, these interactions cross the insulatory TAD boundary, suggesting that their stability might depend on mechanisms other than the standard loop extrusion, possibly involving the chromatin segregation. The pattern of these interactions also varied significantly across different tissues, likely depending on the overall epigenetic activity status of the locus.

Next, we analyzed the expression of six genes within the locus of interest (*Unc5b*, *Slc29a3*, *Psap*, *Vsir*, *Cdh23*, and *Sgpl1*) across five different organs. Out of 60 evaluations, we found 20 instances of significant gene expression alterations linked to deletions at CTCF binding sites. These changes were generally modest, with none exceeding a 50 % variation in expression levels. Notably, the alterations exhibited a tissue-specific pattern, varying in both magnitude and direction. For instance, expression of the *Slc29a3* gene decreased in the kidney but increased in the cerebellum. We confirmed these alterations using digital PCR.

We attempted to fit the observed expression changes in the existing models of interactions between promoters and regulatory elements. Given the tissue-specific nature of these expression changes, we hypothesized that these differences are likely explained by the local epigenetic environment and normal transcriptional activity of the genes in a given tissue.

Following the deletion, the expression profiles of genes divided by the TAD boundary, specifically *Slc29a3* and *Unc5b*, became more similar. In the kidney, *Slc29a3* showed a nearly 40 % decrease in expression, while *Unc5b* was upregulated by 40 %. In wild-type environment, *Slc29a3* is active while *Unc5b* is repressed in this tissue. The disruption of the insulating boundary between these genes appears to cause them to adopt each other's regulatory landscapes, leading to an "averaging" of their expressions.

In the cerebellum, the *Slc29a3* gene exhibited a 15 % increase in expression in a context where *Unc5b* is highly expressed relative to its levels in other tissues. This rise in *Slc29a3* expression might be attributed to its interaction with active cis-regulatory elements from *Unc5b*, such as the super enhancer signatures within Unc5b's first intron. This suggests that the spatial interaction with the *Unc5b* gene allows *Slc29a3* to tap into *Unc5b* active chromatin environment, enhancing its own expression. Unlike in the kidney, there was no corresponding decrease in *Unc5b* expression, likely due to the specific mechanisms of enhancer action or the underlying scale of these changes.

In the cerebellum, we observed a substantial decrease in *Cdh23* expression, which could typically be attributed to either the disconnection of the promoter from essential regulatory elements or the spreading of heterochromatin. Yet, no obvious enhancer elements altering their contact frequency with the *Cdh23* promoter upon TAD disruption were identified. Since Hi-C experiments showed the emergence of ectopic long-range

interactions between the *Cdh23* gene body and the distal regions of the *Unc5b* TAD, we suggest that these new interactions can be associated with the formation of repressive chromatin domains, leading to *Cdh23* downregulation in the cerebellum.

Conclusion: We have shown that the specific deletions of CTCF binding sites in the *Slc29a3*/*Unc5b* locus disrupted the TAD boundary and altered the internal loop architecture within the *Slc29a3* TAD. Although there was no full merging of domains across the organs we studied, the boundary itself was relocated to an inner CTCF-binding site at the promoter of the *Cdh23* gene. Additionally, we observed the formation of new long-range inter-TAD spatial contacts. The origins of these contacts are particularly intriguing as they do not align with any currently understood mechanisms, suggesting novel aspects of chromatin organization and interaction that are yet to be fully elucidated.

In this work, we have developed a sequencing-based method that allows for precise quantification of gene expression changes resulting from TAD reorganization. Accuracy of this technique is compatible with digital PCR, affirming its utility for evaluating the effects of cis-acting mutations. Using this approach, we uncovered an alterations gene expression levels caused by TAD boundary disruption.

The patterns of gene expression changes we observed were not only tissue-specific in terms of magnitude but also in the direction of these changes. Our observations suggest that actively transcribed genes generally experience a decrease in expression levels, while repressed genes tend to gain new enhancer interactions, exemplified by the cases of interactions between *Slc29a3* and *Unc5b* in kidney and cerebellum. Similarly, the *Cdh23* gene, typically highly active in the cerebellum, exhibited a reduction in expression following TAD reorganization. Therefore, the effects of TAD boundary disruptions are expected to vary not only in amplitude but also in direction based on the epigenetic state and differentiation trajectory of the locus.

Funding: The study, including animal work and NGS sequencing, was supported by RSF Project 22-14-00247.

Acknowledgments: NGS data analyses were performed on the nodes of HPC cluster of the Institute of Cytology and Genetics, supported by the project No. 121031800061-7. ChIP-seq and Hi-C assays were performed using equipment of the Novosibirsk State University, supported by the Ministry of Education and Science of Russian Federation, grant # FSUS-2024-001.

doi 10.18699/bgrs2024-1.3-16

The role of the spatial chromatin structure in the regulation of the human keratin gene locus expression

Shtompel A.S.[1, 2]*, Kalabusheva E.P.[3], Luzhin A.V.[1], Molodova M.N.[1, 4], Selivanovsky A.V.[1, 2], Khrameeva E.E.[4], Ulianov S.V.[1, 2], Razin S.V.[1, 2]

[1] *Institute of Gene Biology, RAS, Moscow, Russia*

[2] *Lomonosov Moscow State University, Moscow, Russia*

[3] *Koltzov Institute of Developmental Biology, RAS, Moscow, Russia*

[4] *Skolkovo Institute of Science and Technology, Moscow, Russia*

* *shtompel.an@gmail.com*

Key words: 3D genome; keratins; C-methods; epidermal differentiation

Motivation and Aim: The most notable example of the 3D genome organization influence on transcription is the gene expression switching in multigenic tissue-specific gene clusters. One of them is the keratin gene cluster which is located in chromosome 12 (12q13.13) and contains 26 genes. Potential locus control regions (LCRs) that we have previously discovered could be "switchers" of gene expression and transcriptional regulators of entire multigenic locus. Thus, the aim was to study the spatial structure of the keratin gene locus in mature keratinocytes as well as during the differentiation, including the analysis of LCR contribution in the formation of the locus topology. We also analyzed the mechanisms involved in the formation of the spatial configuration using drug treatments.

Methods and Algorithms: Immortalized normal skin keratinocyte cell line HaCaT, primary skin keratinocytes and iPSC-based epidermal differentiation model were used to analyze keratin gene locus structure. We performed 3C-based chromatin target enrichment method (C-TALE) and FISH coupled with confocal microscopy to obtain data about spatial configuration of the keratin gene locus. ChIP-seq method was used to analyze transcription factors and histone modification profiles in cell lines. We have also treated cells with a panel of small inhibitors to distinguish the effect of transcription and phase separation on the locus structure.

Results: We analyzed spatial chromatin structure of the keratin gene locus during differentiation of induced pluripotent stem cells (iPSC) using C-TALE. We found sequential formation of contacts between LCRs and *KRT5* gene. At the intermediate stage of differentiation contact between LCR1 and LCR2 is established before *KRT5* gene activation. Gene activation is observed on the final day of differentiation accompanied with broad H3 lysine 27 acetylation of *KRT5* promoter as well as LCRs. CTCF and cohesin binding patterns, obtained using ChIP-seq, are present only in LCRs and absent in *KRT5* promoter suggesting that different mechanisms contribute to the contacts formation. FISH method confirms the data obtained by C-protocol and demonstrates the condensation of the keratin gene locus during differentiation. The formation of the transcriptional hub is also indicated by the partial colocalization of LCRs and *KRT5* promoter observed in mature keratinocytes (HaCaT cells) and differentiating cells. We also analyzed the role of transcription and phase separation in formation of specific locus topology in HaCaT and primary skin keratinocytes using drug treatment. Treatment with flavopiridol, inhibiting the transcription initiation of

RNA polymerase II, led to enhancing of existing contacts between *KRT5* and LCRs, but not between LCRs and to the formation of new contacts in gene locus. Conversely treatment with, triptolide, which leads to proteasome-dependent degradation of RNA polymerase II, causes a significant weakening of similar contacts. Treatment with inhibitors of the phase condensate formation – JQ1 and 1,6-hexanediol – shows similar effects as a triptolide, while the contact between LCRs still remains stable.

Conclusion: We suggest that LCRs may play an important role in spatial structure and transcription control within the keratin gene locus during differentiation. Cell treatments with inhibitors of transcription and phase separation, as well as ChIP-seq data indicate the presence of different mechanisms affecting the formation of regulatory interactions in chromatin locus. Contact between LCRs probably forms by cohesion-driven loop extrusion and does not depend on transcriptional status of the locus. Contrariwise KRT5 gene promoter is not bound by cohesin and CTCF, and its contacts are formed by the formation of phase condensates including activator transcription factors and RNA polymerase II.

Funding: The study was supported by the Russian Ministry of Science and Higher Education (075-15-2024-539).

doi 10.18699/bgrs2024-1.3-17

Biop-C: a novel Hi-C based approach for PGT-SR

Stepanchuk Y.K.[1, 2]*, Gridina M.M.[1], Torgunakov N.Yu.[1, 2], Chuyko E.A.[1], Lagunov T.A.[1, 2], Saifitdinova A.F.[3, 4], Nevskaya E.E.[4], Kanbekova O.R.[5], Fishman V.S.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Herzen State Pedagogical University of Russia, St. Petersburg, Russia*

[4] *International Centre for Reproductive Medicine, St. Petersburg, Russia*

[5] *Regional Perinatal Center named after I.D. Yevtushenko, Tomsk, Russia*

** y.stepanchuk@g.nsu.ru*

Key words: Hi-C; single cell Hi-C; chromosomal structural variations; chromosomal aberrations; preimplantation genetic testing

Motivation and Aim: A high incidence of chromosomal abnormalities in early human embryos can affect their viability, implantation, and also cause abortion and congenital diseases [1]. Couples who turn to IVF have the opportunity to test embryos before implantation. Current PGT methods detect aneuploidy, CNV and unbalanced translocations (due to altered copy number of regions included in translocation), but there is no effective approach which is able to differentiate between balanced and normal embryos. Implantation of balanced embryo should result in birth of phenotypically normal child, who will encounter the same reproduction problems. Possibility of differentiation balanced embryos is also very important in case of X-autosomal translocation, because the phenotype of balanced carriers is unpredictable due to random X inactivation [2]. Hi-C can detect all types of chromosomal abnormalities, including balanced translocations. The aim of our work is optimization of Hi-C protocol for preimplantation genetic testing.

Methods and Algorithms: After culturing of human embryos for 5–6 days biopsy for standard PGT-A and vitrification was performed. Based on the patients informed consent, the embryos were thawed and re-biopsied. The trophectoderm biopsy and the embryo were fixed in a 2 % paraformaldehyde solution and then analyzed using the Biop-C protocol (modified protocol for single nuclei [3]).

Results: We evaluated the possibility of Biop-C to detect different types of chromosomal abnormalities in trophectoderm biopsy samples. Using our method, we prepared Hi-C libraries out of 115 embryos and biopsy samples. For 30 samples (10 trophectoderm biopsy samples and 20 embryos) there were sufficient sequencing depth (more than 1 million read pairs) and known PGT status. Ploidy concordance was 90 % for biopsy samples and 85 % for embryos. We were able to detect all CNV associated with unbalanced translocations analyzing both coverages and Hi-C maps. Biop-C analysis also revealed 4 balanced translocations, one of them Robertsonian (Fig. 1), which are consistent with parent's karyotypes.

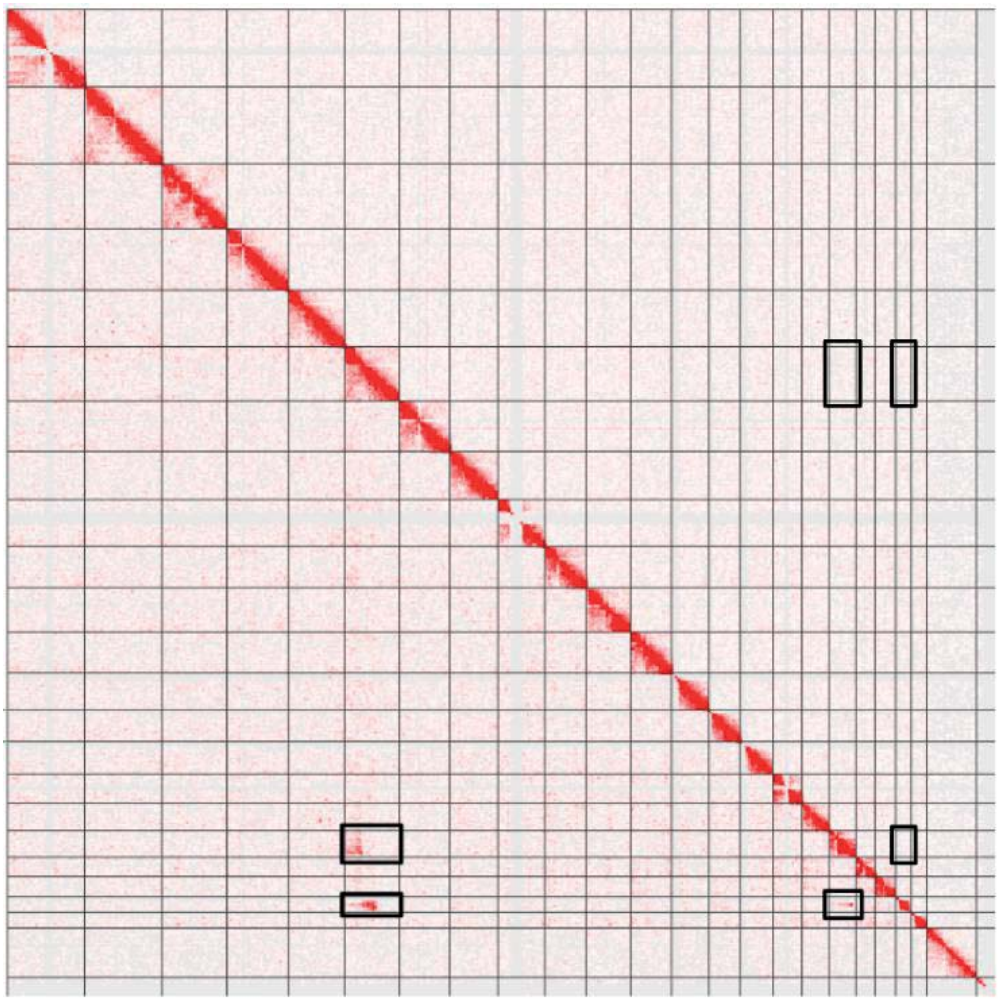


Fig. 1. The heat maps of the Hi-C contacts for embryo Vla1_e below diagonal, above — Sheg1_e as a control. Black rectangles represent unbalanced translocation between chromosomes 6 and 18, and balanced ones between chromosome 6 and 21 and 18 and 21 in Vla1_e sample

Conclusion: The Biop-C protocol allows us to analyze biopsy material of trophectoderm and human embryos and detect aneuploidies, CNVs and translocations, including balanced ones.

Funding: This work was supported by grants from the Russian Science Foundation 22-14-00247.

References

1. Баранов В.С., Кузнецова Т.В. Цитогенетика эмбрионального развития человека: Научно-практические аспекты. СПб: Издательство Н-Л, 2006
2. Ferfouri F., Bernicot I., Schneider A., Haquet E., Hédon B., Anahory T. Is the resulting phenotype of an embryo with balanced X-autosome translocation, obtained by means of preimplantation genetic diagnosis, linked to the X inactivation pattern? *Fertil Steril*. 2016;05(4):1035-1046
3. Gridina M., Taskina A., Lagunov T., Nurislamov A., Kulikova T., Krasikova A., Fishman V. Comparison and critical assessment of single-cell Hi-C protocols. *Heliyon*. 2022;8(10):e11023

doi 10.18699/bgrs2024-1.3-18

Toward analyzing interactions between plasmid DNA and chromatin in HEK293T cells

Yan A.P.[1, 2]*, Salnikov P.A.[1, 2], Gridina M.M.[1, 2], Belokopytova P.S.[1, 2], Fishman V.S.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *a.yan@g.nsu.ru*

Key words: chromatin; plasmid DNA; chromosome conformation capture; 4C; spatial contacts

Motivation and Aim: With the advent of methods for capturing chromosome conformation, studies of interphase nucleus chromatin organization have received a new impetus for development. Using these approaches, elementary units of native chromatin stacking – topologically associated domains (TADs) formed by different mechanisms in vertebrates and invertebrates – were discovered for different animal species [1]. However, how nucleus chromatin interacts with extracellular DNA such as viral and plasmid DNA in space remains poorly understood. This work focuses on developing a method to detect spatial contacts between plasmid DNA and chromatin of eukaryotic cells based on 4C technology.

Methods and Algorithms: We transfect HEK293T cells with a plasmid vector adapted for the 4C experiment. Presumably, in the nucleus, chromatin proteins bind to plasmid DNA and condition its distribution within the nucleus. Treatment of cells with formaldehyde ensures fixation of spatial contacts of genomic and plasmid DNA. We perform the 4C procedure on this material. The resulting 4C library contains predominantly DNA fragments that include the plasmid sequence and the nearby genomic DNA sequence. In this way, the sequences of genome loci that were in contact with the plasmid are revealed.

Results: Optimization of the standard conditions of the 4C method and the design of a special plasmid vector made it possible to efficiently detect plasmid-DNA contacts: over 90 % of 4C library reads contain genomic DNA sequence. The remaining share consists of reads formed as a result of incomplete hydrolysis of plasmid DNA or recovery of the full-length plasmid molecule after ligation. Such products are defined as intramolecular cis contacts and are a characteristic fraction in libraries obtained by chromosome conformation capture methods. We also showed that at least part of the contacts between plasmid and chromatin are not an artifact of the experiment, but are formed within the nucleus. For this purpose, we performed a 4C experiment on a mixture of transfected human cells, mouse cells, and solubilized plasmid. The sequences of the transfected plasmid and the one added to the mixture of cells differed from each other, which allowed us to distinguish the chromatin contacts formed with their participation from each other. The plasmid transfected into cells formed more contacts with human DNA than with mouse DNA, while ligations of the plasmid added to the mixture of cells with DNA of both species occurred with approximately the same frequency (all calculations were performed taking into account the ratio of concentrations of different chromatins in the reaction). Statistically significant differences between the contact patterns of plasmids with DNA from the two species ($p < 0.001$) suggest that about 30 % of the

contacts formed between plasmid transfected into cells and human chromatin are formed through biological patterns within the nucleus rather than by random diffusion in solution.

Conclusion: The proposed 4C-plasmid experiment allows us to detect spatial contacts of plasmid DNA with eukaryotic chromatin. In the future, this approach opens prospects for testing various short DNA sequences cloned into the vector as possible genetic determinants of nucleus compartmentalization and search for methylated loci of the genome.

Funding: The study is supported by RSF project 22-14-00247.

References

1. Rowley M.J., Corces V.G. The three-dimensional genome: principles and roles of long-distance interactions. *Curr Opin Cell Biol*. 2016;40:8-14. doi 10.1016/j.ceb.2016.01.009

1 Симпозиум «Геномика, транскриптомика и биоинформатика»

Symposium “Genomics, transcriptomics and bioinformatics”

doi 10.18699/bgrs2024-1.4-01

Поиск и функциональная аннотация многодоменных белков семейства фосфолипаз А2 у плоских червей

Бочарникова М.[1, 2]*, Турнаев И.[1], Афонников Д.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

** lachynova@bionet.nsc.ru*

Ключевые слова: плоские черви; паразиты человека; фосфолипаза А2; филогения

Мотивация и цель: Семейство ферментов гидролаз – фосфолипазы типа А2 широко известны наличием множества важных биологических функций. Они участвуют в базовых биохимических реакциях человека и большинства других организмов [1]. Впервые ФА2 были обнародованы как цитотоксины яда змей и ферменты панкреатического сока свиней. На данный момент стали широко известны антипаразитические свойства этих ферментов [2]. Также объектом нашего исследования являлись плоские черви. Паразитические плоские черви – один из наиболее распространенных типов гельминтов человека и животных на территории России. А продолжительная болезнь (например, описторхоз) ведет к канцерогенезу [3]. На данный момент есть только разрозненные знания о ФА2 плоских червей. Неизвестно, какие именно функции выполняет ФА2 паразитических плоских червей. Однако есть исследования, где показано, что ФА2 может выступать в качестве маскировки паразита в организме хозяина [4]. Целью работы является поиск генов и аннотация функций белков фосфолипаз А2 у плоских червей.

Методы и алгоритмы: Для поиска генов белков ФА2 у плоских червей мы разработали вычислительный конвейер OrthoDom, который включает два основных этапа: фильтрация специфических доменов и кластеризация ортологов исследуемых организмов. Мы исследовали белковые последовательности плоских червей 44 видов – 2 свободноживущих и 42 паразитических. Для идентификации белков фосфолипаз А2 у плоских червей на разные типы мы использовали референсные последовательности фосфолипаз позвоночных, которые были классифицированы по 11 основным типам. Мы определяли принадлежность белковых последовательностей червей к фосфолипазам определенного типа, если они попали в одну ортогруппу с референсными белками фосфолипаз позвоночных.

Особенностью алгоритма Orthofinder является поиск ортологов на основе быстрого парного сравнения последовательностей, которое в случае многодоменных белков (к которым относятся фосфолипазы) может давать ошибки. Чтобы учесть разнообразие доменного состава фосфолипаз, мы проводим дополнительную верификацию последовательностей, принадлежащих к ортогруппам, включающим фосфолипазы позвоночных: если набор доменов, характерных для данного типа фосфолипаз, у белков червей отличается, мы исключали такие белки из списка. Также был проведен анализ экспрессии ФА2 у паразитических и свободноживущих плоских червей.

Результаты: Главным результатом нашего исследования является корреляция наличия ФА2 различных типов с образом жизни плоских червей (свободноживущие и паразиты). У паразитических плоских червей наблюдается редукция большинства типов ФА2. Однако у присутствующих типов ФА2 высокая дупликация этих генов в геномах. У свободноживущих червей, наоборот, представлено большинство типов ФА2, но в основном единоразово.
Выводы: Эти результаты позволяют лучше понять связь между разнообразием ФА2 и специфическим образом жизни плоских червей, что может иметь важные иммунологические и эволюционные последствия.
Финансирование: Исследование поддержано Курчатовским геномным центром ИЦиГ СО РАН (проект № 075-15-2019-1662): разработка метода идентификации ортологов OrthoDom.

Search and functional annotation of multi-domain proteins from the phospholipase A2 family in flatworms

Bocharnikova M.[1, 2]*, Turnaev I.[1], Afonnikov D.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *lachynova@bionet.nsc.ru*

Key words: flatworms; human parasites; phospholipase A2; phylogeny

Motivation and Aim: The family of hydrolase enzymes, phosphopolypases type A2, is widely known for having many important biological functions. They are involved in the basic biochemical reactions of humans and most other organisms [1]. For the first time, FA2 was made public as cytotoxins of snake venom and enzymes of pig pancreatic juice. At the moment, the antiparasitic properties of these enzymes have become widely known [2]. Flatworms were also the object of our research. Parasitic flatworms are one of the most common types of helminths of humans and animals in Russia. And prolonged illness (for example, opisthorchiasis) leads to carcinogenesis [3]. At the moment, there is only scattered knowledge about FA2 flatworms. It is unknown exactly what functions the FA2 of parasitic flatworms performs. However, there are studies that have shown that FA2 can act as a parasite disguise in the host body [4]. The aim of the work is to search for genes and abstract the functions of phospholipase A2 proteins in flatworms.
Methods and Algorithms: To search for FA2 protein genes in flatworms, we have developed the OrthoDom computational pipeline, which includes the following two main stages – filtering of specific domains and clustering of orthologs of the studied organisms. We studied the protein sequences of flatworms of 44 species, 2 free-living and 42 parasitic. To identify phospholipase A2 proteins in flatworms into different types, we used reference sequences of vertebrate phospholipases, which were classified into 11 main types. We determined the belonging of worm protein sequences to phospholipases of a certain type if they fell into the same orthogroup with reference proteins of vertebrate phospholipases.
A feature of the Orthofinder algorithm is the search for orthologs based on fast paired sequence comparison, which in the case of multi-domain proteins (which include phospholipases) can give errors. To take into account the diversity of the domain composition of phospholipases, we perform additional verification of sequences

belonging to orthogroups including vertebrate phospholipases: if the set of domains characteristic of this type of phospholipase in worm proteins differs, we excluded such proteins from the list. FA2 expression was also analyzed in parasitic and free-living flatworms.

Results: These results provide a better understanding of the relationship between FA2 diversity and the specific lifestyle of flatworms, which may have important immunological and evolutionary implications.

Funding: The study was supported by the Kurchatov Genomic Center of the ICG SB RAS (project No. 075-15-2019-1662): development of the OrthoDom method for orthologs identification.

Список литературы/References

1. Dennis E.A., Cao J., Hsu Y.H., Magrioti V., Kokotos G. Phospholipase A2 enzymes: physical structure, biological function, disease implication, chemical inhibition, and therapeutic intervention. *Chem Rev*. 2011;111(10):6130-6185. doi 10.1021/cr200085w
2. Handbook of natural toxins. V. 3. Marine toxins and venoms. New York and Basel, 1988
3. Scholte L.L.S., Pascoal-Xavier M.A., Nahum L.A. Helminths and cancers from the evolutionary perspective. *Front Med*. 2018;5:90
4. Teixeira S.C., da Silva M.S., Gomes A.A.S., Moretti N.S., Lopes D.S., Ferro E.A.V., Rodrigues V.M. Panacea within a Pandora's box: the antiparasitic effects of phospholipases A_2 (PLA_2s) from snake venoms. *Trends Parasitol*. 2022;(1):80-94. doi 10.1016/j.pt.2021.07.004

doi 10.18699/bgrs2024-1.4-02

Анализ событий альтернативного сплайсинга в децидуальных клетках плаценты человека

Гавриленко М.М.*, Трифонова Е.А., Бабовская А.А., Зарубин А.А., Степанов В.А.
Научно-исследовательский институт медицинской генетики Томского национального исследовательского медицинского центра РАН, Томск, Россия
** mmgavrilenko@gmail.com*

Ключевые слова: альтернативный сплайсинг; полнотранскриптомное секвенирование; транскрипт; плацента; децидуальные клетки; MAJIQ; rMATS; SGSeq

Мотивация и цель: Изучение альтернативного сплайсинга (АС) представляет значительный интерес, поскольку данный процесс является фундаментальным в молекулярной биологии, в ходе которого из одной пре-мРНК может быть получено множество транскриптов мРНК, что приводит к экспрессии различных изоформ белка из одного гена. Известно, что АС регулирует экспрессию генов, участвующих в развитии и функционировании плаценты. Изменение экспрессии плацентарных генов может нарушить такие важные процессы, как дифференцировка трофобласта, ангиогенез, транспорт питательных веществ и иммунный гомеостаз, что может привести к осложнениям беременности. Децидуальные клетки (ДК) являются основной клеточной популяцией материнской части плаценты и играют ключевую роль в поддержании физиологической беременности [1]. Немаловажным является выбор биоинформатического инструмента для анализа событий АС, которых достаточно много на сегодняшний день. Методы анализа, основанные на событийном подходе, обладают более высокой чувствительностью и специфичностью даже при низких уровнях экспрессии. В данном случае обычно используются такие инструменты, как MAJIQ [2], rMATS [3], SGSeq [4].
Методы и алгоритмы: Глубокое секвенирование проведено для 8 образцов плацентарной ткани, полученной от женщин с физиологическим течением беременности. Расчетное число прочтений на образец составляло не менее 60 млн. Анализ качества полученных сиквенсов выполнен с помощью программы FastQC. Последовательности адаптеров удалены с помощью программы Trimmomatic. Выравнивание полученных транскриптомных данных на референсный геном человека (GRCh38) выполнено с помощью программы STAR. Далее для анализа событий АС выбраны программы MAJIQ, rMATS, SGSeq, так как они существенно обновлены в 2023–2024 гг. и имеют возможность визуализации.
Результаты: С помощью программы MAJIQ идентифицировано 3501 событие АС для 2731 гена; с помощью программы rMATS идентифицировано 66 687 событий для 14 784 генов; с помощью программы SGSeq – 15 782 события для 5616 генов. При анализе результатов полученных данных важно отметить, что количество генов, подверженных АС, значительно варьировало. Идентифицировано 15 887 уникальных плацентарных генов, подверженных альтернативному сплайсингу. При поиске общих генов показана низкая репликация результатов (33.9 %), так как всего 5387 генов выявлены хотя бы в двух программах одновременно. Низкая репликация и высокая вариативность в количестве идентифицированных событий

и генов могут быть связаны с различными статистическими и алгоритмическими подходами, заложенными разработчиками в программы для анализа альтернативного сплайсинга.

Следующим шагом все события были классифицированы согласно 7 типам событий АС для программ MAJIQ и SGSeq (пропуск экзона, альтернативные 5' или 3' сайты сплайсинга, удержание интрона, альтернативные первый и последний экзоны и взаимоисключающие экзоны) и 5 типам событий для программы rMATS (без альтернативных первого и последнего экзонов). Наиболее распространенным типом АС является пропуск экзона (46.4 %), что показано для всех трех программ и согласуется с давно известным положением о преобладании пропуска экзона у млекопитающих. Далее наблюдается вариативность. В программе MAJIQ альтернативный первый экзон является вторым по популярности событием АС (24.8 %), в rMATS – альтернативный 3'-сайт сплайсинга (14.1 %), тогда как в SGSeq – удержание интрона (28.6 %), что подтверждается новыми данными о росте событий удержания интрона [5]. Наименее редким событием являются взаимоисключающие экзоны (1.3 %), что также показано для всех трех программ.

В целях более детальной функциональной аннотации генов, подверженных АС, был проведен анализ 1857 общих для всех трех программ генов. Анализ реконструированной с использованием программы STRING белок-белковой сети включал в себя 386 белковых продуктов (high confidence = 0.999) и 98 кластеров, из которых 8 кластеров включали в себя 8 генов и более, 57 кластеров по 2 гена, количество генов в остальных кластерах варьировало от 3 до 5 включительно. Далее проведена подробная аннотация первых по количеству генов 3 кластеров в базах данных GeneOntology, KEGG, Reactome с помощью онлайн-ресурса WebGestalt (https://www.webgestalt.org/).

Кластер 1 включает 26 генов, центральное место с максимальным числом и силой взаимодействий (node_degree ≥ 10, score ≥ 0,99) занимает ген *ACTL6A*. Согласно результатам функциональной аннотации, сигнальные пути кластера 1 преимущественно связаны с модификацией хроматина и регуляцией транскрипции. Во время беременности изменения хроматина играют важную роль в регуляции экспрессии генов в развивающемся плоде, плаценте и тканях матери. Одним из примеров модификации хроматина во время беременности является метилирование ДНК. Известно, что изменения в метилировании материнской ДНК могут привести к гестационному диабету, анемии, преждевременным родам и преэклампсии [6].

Кластер 2 включает 17 генов, центральное место с максимальным числом и силой взаимодействий (node_degree ≥ 8, score ≥ 0.99) занимает ген *SRC*. Согласно результатам функциональной аннотации, сигнальные пути кластера 2 преимущественно связаны с межклеточным сигналами, миграцией клеток и межклеточной адгезией. Молекулы клеточной адгезии играют решающую роль в обеспечении адгезии и миграции клеток трофобласта в эндометрий и последующей инвазии в кровеносные сосуды матери [7]. Нарушение регуляции этих процессов адгезии может способствовать развитию таких состояний, как преэклампсия или задержка роста плода [8].

Кластер 3 включает 15 генов, центральное место с максимальным числом и силой взаимодействий (node_degree ≥ 5, score ≥ 0.99) занимают гены *TBK1*, *TRAF3*. Согласно результатам функциональной аннотации, сигнальные пути кластера 3 преимущественно связаны с иммунной системой, апоптозом и сигнальным путем

NF-kappaB (ядерный фактор каппа B), который является ключевым фактором транскрипции, участвующим в регуляции иммунных и воспалительных реакций, пролиферации, дифференцировке и выживании клеток. Адекватная регуляция активности NF-κB имеет решающее значение для поддержания иммунного гомеостаза матери и плода и обеспечения успешного исхода беременности. Апоптоз также способствует поддержанию иммунной толерантности на уровне взаимодействия матери и плода, устраняя активированные иммунные клетки и предотвращая чрезмерные воспалительные реакции, которые могут нанести вред развивающемуся плоду. Нарушение регуляции этих процессов может способствовать возникновению самопроизвольного аборта и преждевременных родов [9].
Выводы: В исследовании проведена оценка событий АС в децидуальных клетках плаценты при физиологической беременности с помощью трех биоинформатических программ MAJIQ, rMATS, SGSeq. Было идентифицировано разное количество генов и событий АС каждой из программ, а также продемонстрирована низкая репликация результатов, что может быть связано с различными математическими подходами, заложенными разработчиками в алгоритмы программ, в связи с чем целесообразно использовать несколько программ для идентификации событий АС и для последующего анализа выбирать те гены, которые являются общими для всех примененных в исследовании программ. Во всех трех программах наиболее часто встречающимся из выявленных событий является пропуск экзона, а самым редким – взаимоисключающий пропуск экзонов. Анализ реконструированной генной сети позволил выявить биологические пути, ассоциированные с иммунной системой, клеточной миграцией, межклеточными контактами и регуляцией экспрессии. Следует отметить, что ряд генов, подверженных АС в ДК, ассоциирован с осложнениями беременности. Полученные результаты подтверждают важность альтернативного сплайсинга, который существенно увеличивает транскрипционное разнообразие и представляет собой значимый механизм регуляции генов в децидуальных клетках.
Финансирование: Исследование выполнено за счет средств государственного задания по теме ФНИ № 122020200083-8.

Analysis of alternative splicing events in human placental decidual cells

Gavrilenko M.M.*, Trifonova E.A., Babovskaya A.A., Zarubin A.A., Stepanov V.A.
Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia
** mmgavrilenko@gmail.com*

Key words: alternative splicing; whole transcriptome sequencing; transcript; placenta; decidual cells; MAJIQ; rMATS; SGSeq

Motivation and Aim: The alternative splicing (AS) study is of considerable interest, because this process is fundamental in molecular biology. Multiple mRNA transcripts can be obtained from a single pre-mRNA during AS. It leads to the expression of various protein isoforms from a single gene. It is known that AS regulates the expression of genes involved in the development and functioning of the placenta. Changes in the placental genes expression can disrupt important processes such as trophoblast

differentiation, angiogenesis, nutrient transport and immune homeostasis, which can lead to pregnancy complications. Decidual cells (DC) are the main cell population of the placenta maternal part and have a key role in maintaining physiological pregnancy [1]. It is also important to choose a bioinformatics tool for analyzing AS events among a large number of them. Event-based analysis methods have higher sensitivity and specificity even at low expression levels. In this case, tools such as MAJIQ [2], rMATS [3], SGSeq [4] are usually used.

Methods and Algorithms: Deep sequencing was performed for 8 placental tissue samples obtained from women with a physiological pregnancy. The calculated number of readings per sample was at least 60 million. The quality analysis of the received transcriptomic data was performed using the FastQC. The adapter sequences are deleted using the Trimmomatic. The alignment to the human reference genome (GRCh38) was performed using the STAR. The MAJIQ, rMATS and SGSeq programs were selected for the analysis of AS events, because they have the ability to visualize and were significantly updated in 2023–2024.

Results: As a result of the study, 3501 AS events for 2731 genes were identified with MAJIQ; 66687 events for 14784 genes were identified with rMATS; 15782 events for 5616 genes were identified with SGSeq. It is important to note that the number of genes affected by AS varied significantly. 15887 unique placental genes subject to alternative splicing have been identified. Low replication of the results was shown (33.9 %), because only 5387 genes were identified in at least two programs. Low replication and high variability in the number of identified events and genes may be associated with various statistical and algorithmic approaches embedded by developers in programs for analyzing AS.

All events were classified according to 7 types of AS events for the MAJIQ and SGSeq programs (exon skipping, alternative 5' or 3' splicing sites, intron retention, alternative first and last exons and mutually exclusive exons) and 5 types of events for the rMATS program (without alternative first and last exons). The most common type of AS is exon skipping (46.4 %), which is shown for all three programs. It is consistent with the long-known statement on the predominance of exon skipping in mammals. Further, there is variability. The alternative first exon is the second most popular AS event (24.8 %) for MAJIQ, the alternative 3'-splicing site (14.1 %) for rMATS and intron retention (28.6 %) for SGSeq. It is confirmed by new data on the growth of intron retention events [5]. The least rare event is mutually exclusive exons (1.3 %), which is also shown for all three programs.

In order to provide a more detailed functional annotation of genes affected by AS, 1857 genes common to all three programs were analyzed. The analysis of the protein-protein network reconstructed using the STRING program included 386 protein products (high confidence = 0.999) and 98 clusters, of which 8 clusters included 8 or more genes, 57 clusters consist of 2 genes, the genes number in the other clusters varied from 3 to 5 genes inclusive. Next, a detailed annotation of the first 3 clusters in terms of the genes number in the GeneOntology, KEGG, Reactome databases was carried out using the WebGestalt online resource (https://www.webgestalt.org/).

Cluster 1 includes 26 genes. The *ACTL6A* gene is the central gene with the maximum number and strength of interactions (node_degree $\geq$ 10, score $\geq$ 0.99). According to the results of the functional annotation, the cluster 1 signaling pathways are mainly associated with chromatin modification and transcription regulation. During pregnancy,

chromatin changes play an important role in regulating gene expression in the developing fetus, placenta, and maternal tissues. One example of chromatin modification during pregnancy is DNA methylation. It is known that changes in maternal DNA methylation can lead to gestational diabetes, anemia, premature birth and preeclampsia [6].

Cluster 2 includes 17 genes. The *SRC* gene is the central gene with the maximum number and strength of interactions (node_degree ≥ 8, score ≥ 0.99). According to the results of the functional annotation, the cluster 2 signaling pathways are mainly associated with intercellular signals, cell migration and intercellular adhesion. Cell adhesion molecules play a crucial role in providing adhesion and trophoblast cells migration into the endometrium and subsequent invasion into the mother blood vessels [7]. A disorder of these adhesion processes regulation can contribute to the preeclampsia and fetal growth restriction development [8].

Cluster 3 includes 15 genes. The *TBK1* and *TRAF3* are the central genes with the maximum number and strength of interactions (node_degree ≥ 5, score ≥ 0.99). According to the results of the functional annotation, cluster 3 signaling pathways are predominantly associated with the immune system, apoptosis and the NF-kappaB signaling pathway, which is a key transcription factor involved in the regulation of immune and inflammatory reactions, cell proliferation, differentiation and survival. Adequate regulation of NF-kappaB activity is crucial for maintaining maternal and fetal immune homeostasis and provides a successful pregnancy outcome. Apoptosis also helps maintain immune tolerance at the level of maternal-fetal interaction by eliminating activated immune cells and preventing excessive inflammatory reactions that can harm the developing fetus. A violation these biological processes can contribute to the spontaneous abortion and premature birth [9].

Conclusion: In the study was performed assessment of AS events in placental decidual cells during physiological pregnancy using three bioinformatics programs MAJIQ, rMATS, SGSeq. A different number of AS genes and events were identified for each programs. Also low replication of the results was demonstrated. There are may be due to different mathematical approaches embedded by the developers in the programs algorithms. Therefore it is advisable to use several programs to identify AS events and for subsequent analysis select those genes that are common to all tools. In all three programs, the most common of the identified events is an exon skip, and the rarest is a mutually exclusive exons. The reconstructed gene network analysis revealed biological pathways associated with the immune system, cell migration, intercellular contacts and regulation of gene expression. It should be noted that a number of genes susceptible to AS in DC are associated with pregnancy complications. The results obtained confirm the importance of alternative splicing, which significantly increases transcriptional diversity and represents a significant mechanism for gene regulation in decidual cells.

Funding: The research was carried out at the expense of the state assignment (fundamental scientific research No. 122020200083-8).

Список литературы/References

1. Schatz F., Guzeloglu-Kayisli O., Arlier S. Kayisli U.A., Lockwood C.J. The role of decidual cells in uterine hemostasis, menstruation, inflammation, adverse pregnancy outcomes and abnormal uterine bleeding. *Hum Reprod Update*. 2016;4:497-515. doi 10.1093/humupd/dmw004
2. Vaquero-Garcia J., Barrera A., Gazzara M.R. et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife*. 2016;5:e11752. doi 10.7554/elife.11752
3. Shen S., Park J. W., Lu Z.X. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA*. 2014;111(51):E5593-E5601. doi 10.1073/pnas.1419161111

4. Goldstein L.D., Ca Y., Pau G. et al. Prediction and quantification of splice events from RNA-seq data. *PloS One*. 2016;11(5):e0156132. doi 10.1371/journal.pone.0156132
5. Monteuuis G., Wong J.J., Bailey C.G. et al. The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res*. 2019;47(22):11497-11513. doi 10.1093/nar/gkz1068
6. Das J., Maitra A. Maternal DNA methylation during pregnancy: a review. *Reprod Sci*. 2021;28(10):2758-2769. doi 10.1007/s43032-020-00456-4
7. Hohn H.P., Denker H.W. Experimental modulation of cell-cell adhesion, invasiveness and differentiation in trophoblast cells. *Cells Tissues Organs*. 2002:172(3):218-236. doi 10.1159/000066965
8. Barbitoff Y.A., Tsarev A.A., Vashukova E.S. et al. A data-driven review of the genetic factors of pregnancy complications. *Int J Mol Sci*. 2020:21(9):3384. doi 10.3390/ijms21093384
9. Sharp A.N., Heazell A.E., Crocker I.P. et al. Placental apoptosis in health and disease. *Am J Reprod Immunol*. 2010;64(3):159-169. doi 10.1111/j.1600-0897.2010.00837.x

doi 10.18699/bgrs2024-1.4-03

Транскриптомный ответ *Drosophila melanogaster* на инфицирование *Wolbachia*

Грунтенко Н.*, Дерюженко М., Шишкина О., Карпова Е., Бобровских М., Андреенкова О., Бурдина Е., Васильев Г., Шацкая Н.

Институт цитологии и генетики СО РАН, Новосибирск, Россия

** nataly@bionet.nsc.ru*

Ключевые слова: *Wolbachia; Drosophila*; симбиоз; дифференциальная экспрессия генов; углеводно-жировой метаболизм; щелочная фосфатаза; устойчивость к голоданию

Транскриптом клеток дрозофилы, как и любого другого живого организма, тканеспецифичен, весьма лабилен и изменяется под влиянием самых разнообразных биотических и абиотических факторов. Мы исследовали эффект одного из таких факторов – инфицирования эндосимбиотической бактерией *Wolbachia*, постаравшись унифицировать все условия и влияние прочих факторов при сборе инфицированных и контрольных (лишенных инфекции) образцов.

Wolbachia широко известна своей способностью манипулировать репродуктивной системой насекомого-хозяина, успешно распространяясь в его популяциях, однако информация о том, какую выгоду этот симбиоз может приносить хозяину, до сих достаточно фрагментарна и противоречива. Мы продемонстрировали, что у *Drosophila melanogaster* бактерия оказывает существенное влияние на экспрессию генов, задействованных в эмбриогенезе, окислительно-востановительных процессах, пищеварении, транспорте и метаболизме углеводов, а также на экспрессию нескольких генов щелочных фосфатаз. Эти изменения сопровождались увеличением массы тела мух, содержания глюкозы, общих липидов и триглицеридов, коррелирующими с повышенной устойчивостью инфицированных мух к голоданию, что в значительной мере отвечает на вопрос о пользе симбиоза с *Wolbachia* для *D. melanogaster*.

Также важно отметить, что, несмотря на незначительное повышение уровня экспрессии генов щелочных фосфатаз у инфицированных мух, активность соответствующего фермента у них тоже была достоверно повышена по сравнению с контролем, что свидетельствует о том, что даже незначительные изменения в транскриптоме могут иметь значительный биохимический физиологический эффект, так как щелочные фосфатазы участвуют в метаболизме катехоламинов на стадии синтеза тирозина, предшественника нейрогормонов и нейромедиаторов дофамина и октопамина, из тирозин-О-фосфата, и играют важную роль в контроле онтогенеза, поведения, стресс-ответа и приспособленности насекомых.

Финансирование: Исследование поддержано РНФ (№ 21-14-00090).

Drosophila melanogaster transcriptome response to *Wolbachia* infection

Gruntenko N.*, Deryuzhenko M., Shishkina O., Karpova E., Bobrovskikh M., Andreenkova O., Burdina E., Vasiliev G., Shatskaya N.

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

** nataly@bionet.nsc.ru*

Key words: *Wolbachia*; *Drosophila*; symbiosis; differential gene expression; carbohydrate-fat metabolism; alkaline phosphatase; starvation resistance

The transcriptome of *Drosophila* cells, like that of any other living organism, is tissue specific, very labile, and changes under the influence of a wide variety of biotic and abiotic factors. We investigated the effect of one of these factors – infection with the endosymbiotic bacterium *Wolbachia*, trying to standardize all conditions and the influence of other factors when collecting infected and control (non-infected) samples.

Wolbachia is widely known for its ability to manipulate the host insect's reproduction, successfully spreading through host's populations; however, information about possible usefulness of this symbiosis to the host is still quite fragmentary and contradictory. We demonstrate that in *Drosophila melanogaster* the bacterium has significant effects on the expression of genes involved in embryogenesis, oxidation–reduction processes, digestion, carbohydrate transport and metabolism, as well as on the expression of several alkaline phosphatase genes. These changes are accompanied by an increase in fly body weight and contents of glucose, triglycerides and total lipids, correlating with an increased resistance of infected flies to starvation, which largely answers the question of the benefits of symbiosis with *Wolbachia* for *D. melanogaster*.

It is noteworthy that, despite a slight increase in the expression level of alkaline phosphatase genes in infected flies, the activity of the corresponding enzyme in them was also significantly increased compared to the control, which indicates that even minor changes in the transcriptome can have a significant biochemical physiological effect: alkaline phosphatases are involved in the metabolism of catecholamines at the stage of synthesis of tyrosine (a precursor of neurohormones and neurotransmitters dopamine and octopamine) from tyrosine-O-phosphate, and play an important role in the control of ontogenesis, behavior, stress response and fitness in insects.

Funding: The study was supported by RSF grant (No. 21-14-00090).

doi 10.18699/bgrs2024-1.4-04

Поиск и идентификация дифференциально экспрессирующихся генов ответа на тепловой стресс *Drosophila melanogaster* с низкой степенью изменения экспрессии

Дерюженко М.А.*, Грунтенко Н.Е.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
** maksimd@bionet.nsc.ru*

Ключевые слова: транскриптом; *Drosophila melanogaster*; реконструкция генной сети; потеря информации

Мотивация и цель: Потеря информации – это одна из ключевых проблем современного анализа транскриптомных данных. Основной причиной ее возникновения является влияние множества неучтенных факторов, возникающих при проведении и анализе результатов эксперимента, которые снижают общую достоверность полученных результатов. В попытке избавиться от них, как правило, отсекаются дифференциально экспрессирующиеся гены (ДЭГ) с высоким значением коэффициента ложного обнаружения (FDR). При этом такие гены могут играть важную роль в контроле наблюдаемых изменений функций организма и получить высокие значения FDR ввиду низкой степени изменения экспрессии.

Эту проблему можно решить несколькими способами:

1) провести функциональную аннотацию каждого ДЭГ и изучить литературные данные по ним;

2) провести метаанализ транскриптомных данных и оценить встречаемость каждого ДЭГ в различных экспериментах.

Оба варианта имеют существенные недостатки: в первом случае анализ будет занимать большой промежуток времени, а во втором необходимо проанализировать большое количество экспериментов, проведенных в сходных условиях, что не представляется возможным для большинства организмов в связи с ограниченным количеством опубликованных транскриптомных данных. Целью нашей работы является оптимизация второго способа и разработка метода поиска и идентификации значимых дифференциально экспрессирующихся генов с низкой степенью изменения экспрессии.

Мы решили остановиться на классическом объекте генетических исследований – *Drosophila melanogaster*. Ранее на *Drosophila* было доказано наличие единого ответа на различные стрессовые факторы у насекомых и описан его нейрогормональный механизм, отличающийся высокой консервативностью [1, 2]. Нами была выбрана высокая температура, как один из классических способов стрессорного воздействия на организм насекомого. Для уменьшения количества факторов, влияющих на результаты эксперимента, были выдвинуты следующие требования к данным для анализа: библиотеки получены из имаго *Drosophila melanogaster,* которые подвергались стрессовому воздействию в 37 °C в течение 5 минут, после чего были зафиксированы в жидком азоте.

Методы и алгоритмы: Метаанализ транскриптомных данных на основе программы InterTransViewer. Функциональная аннотация генов с использованием баз данных NCBI, Ensembl, FlyBase и утилиты ShinyGO [3]. Реконструкция генной сети с использованием базы данных StringDB.
Результаты: В результате поиска транскриптомных экспериментов, посвященных тепловому стрессу, было найдено всего три работы, две из которых удовлетворяли поставленному условию. В отобранных экспериментах использовались целые самки, что также увеличило степень схожести экспериментальных условий. Ввиду низкого числа экспериментов было решено отказаться от использования формулы расчета весов, предлагаемой программой InterTransViewer [4] и сосредоточиться на дифференциально экспрессирующихся генах (ДЭГ) с низким уровнем достоверности (FDR > 0.05). За счет функциональной аннотации выбранных генов, экспериментальной верификации и дополнительного анализа литературы удалось расширить существующую генную сеть ответа на тепловой стресс и подтвердить работоспособность предложенного метода.
Финансирование: Исследование поддержано бюджетным проектом FWNR-2022-0019 Министерства науки и высшего образования Российской Федерации.

Search and identification of differentially expressed *Drosophila melanogaster* heat stress response genes with a low degree of expression change

Deryuzhenko M.A.*, Gruntenko N.E.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** maksimd@bionet.nsc.ru*

Key words: transcriptome; *Drosophila melanogaster*; gene network reconstruction; information loss

Motivation and Aim: Information loss is one of the key problems of modern transcriptome data analysis. The main reason for its occurrence is the influence of many unaccounted factors arising during the conduct and analysis of experimental results, which reduce the overall reliability of the obtained results. In an attempt to get rid of them, as a rule, differentially expressed genes (DEGs) with high false discovery rate (FDR) are cut off. At the same time, such genes may play an important role in controlling the observed changes in body functions and get high FDR values due to the low degree of expression change.
This problem can be addressed in several ways:
1) Conduct functional annotation of each DEG and review the literature data on them.
2) Conduct a meta-analysis of transcriptomic data and estimate the occurrence of each DEG in different experiments.
Both of these options have significant disadvantages: in the first case, the analysis will take a long period of time, and in the second case it is necessary to analyze a large number of experiments performed under similar conditions, which is not possible for most organisms due to the limited amount of published transcriptomic data. The aim of this work is to optimize the second method and to develop a method for searching and

identifying significant differentially expressed genes with a low degree of expression change.

We decided to focus on a classical object of genetic research – *Drosophila melanogaster*. Previously, a single response to various stress factors in insects was proved in Drosophila and its neurohormonal mechanism characterized by high conservativity was described [1, 2]. We chose high temperature as one of the classical ways of stressor influence on the insect organism. To reduce the number of factors affecting the results of the experiment, the following data requirements for analysis were made: libraries were obtained from *Drosophila melanogaster* adults that were stressed at 37 °C for 5 min, after which they were fixed in liquid nitrogen.

Methods and Algorithms: Meta-analysis of transcriptome data, based on InterTransViewer program. Functional annotation of genes using NCBI, Ensembl, FlyBase databases and the ShinyGO utility [3]. Gene network reconstruction using StringDB database.

Results: The search for transcriptome experiments devoted to heat stress found only three papers, two of which fulfilled the condition set. Also, whole females were used in the selected experiments, which also increased the degree of similarity of experimental conditions. Due to the low number of experiments, it was decided not to use the formula for calculating weights proposed by the InterTransViewer program [4] and to focus on differentially expressed genes (DEGs) with a low level of confidence (FDR > 0.05). Due to functional annotation of the selected genes, experimental verification and additional literature analysis, it was possible to expand the existing gene network of response to heat stress and confirm the performance of the proposed method.

Funding: The study is supported by BP #FWNR-2022-0019 of the Ministry of Science and Higher Education of the Russian Federation.

Список литературы/References

1. Bobrovskikh M.A., Gruntenko N.E. Mechanisms of neuroendocrine stress response in drosophila and its effect on carbohydrate and lipid metabolism. *Insects*. 2023;14(5):474. doi 10.3390/insects14050474
2. Rauschenbach I.Y. et al. Stress-like reaction of Drosophila to adverse environmental factors. *J Comp Physiol B Biochem Syst Environ Physiol*. 1987;157(4):519-531
3. Ge S.X., Jung D., Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*. 2020;36(8):2628-2629. doi 10.1093/bioinformatics/btz931
4. Tyapkin A.V., Lavrekha V.V., Ubogoeva E.V., Oshchepkov D.Y., Omelyanchuk N.A., Zemlyanskaya E.V. InterTransViewer: a comparative description of differential gene expression profiles from different experiments. *Vavilov J Genet Breed*. 2023;27(8):1042-1052. doi 10.18699/VJGB-23-119

doi 10.18699/bgrs2024-1.4-05

Потенциальные регуляторные днРНК и мРНК-мишени опухоль-супрессорной микроРНК miR-124-3p в транскриптоме рака яичников

Зайченко Д.М.[1], Астафьева Я.Р.[1], Урошлев Л.А.[2], Фридман М.В.[2], Бурденный А.М.[1], Логинов В.И.[1], Московцев А.А.[1]*, Брага Э.А.[1]

[1] *Институт общей патологии и патофизиологии, Москва, Россия*

[2] *Институт общей генетики им. Вавилова РАН, Москва, Россия*

* *bioinf@mail.ru*

Ключевые слова: транскриптом; днРНК; SKOV3; экспрессия miR-124-3p; микрочипы Affymetrix HTA 2.0

Мотивация и цель: Мотивом для работы являются идентификация и характеристика регуляторных длинных некодирующих РНК (днРНК) и мРНК-мишеней микроРНК miR-124-3p при раке яичников (РЯ) с использованием транскриптомных данных. Цель – определить функциональный ответ этих РНК на сверхэкспрессию miR-124-3p и идентифицировать потенциальные маркеры РЯ.

Методы и алгоритмы: В работе использовались анализ транскриптомных данных из базы данных NCBI GEO для скрининга регуляторных днРНК и мРНК-мишеней miR-124-3p при раке яичников, корреляционный анализ между miR-124-3p и днРНК/мРНК в транскриптоме рака яичников, трансфекция и сверхэкспрессия в клетках SKOV3 миметиков miR-124-3p, анализ дифференциальной экспрессии днРНК и мРНК в этих клетках с использованием микрочипов Affymetrix HTA 2.0, выравнивание последовательностей между miR-124-3p и выбранными днРНК и мРНК с использованием алгоритма Смита–Уотермана, анализ функционального ответа выбранных днРНК и мРНК-мишеней на сверхэкспрессию miR-124-3p в клетках SKOV3, биоинформатический прогноз потенциальных мишеней мРНК и днРНК, которые взаимодействуют с miR-124-3p.

Результаты: Было показано, что miR-124-3p является критическим опухолевым супрессором при многих типах рака. Скрининг регуляторных днРНК и мРНК-мишеней этой микроРНК при РЯ проводился с использованием набора транскриптомных данных 131 пациента из NCBI GEO (GSE211669). Были идентифицированы наборы из 262 днРНК ($Rs < –0.5$, $p < 0.01$) и 333 мРНК ($Rs < 0.7$, $p < 0.0001$), отрицательно коррелирующих с miR-124-3p. Данные корреляционного анализа РНК трех типов в транскриптоме РЯ сравнивали с функциональным ответом клеток SKOV3 на сверхэкспрессию miR-124-3p. Мы трансфицировали SKOV3 с использованием имитатора miR-124-3p и проанализировали дифференциально экспрессируемые (ДЭ) днРНК и мРНК с использованием высокопроизводительных микрочипов Affymetrix HTA 2.0. Среди 333 мРНК, отрицательно коррелирующих с miR-124, в клетках SKOV3 было обнаружено 127 ДЭ мРНК, из которых 94 мРНК были подавлены в 1.1–1.7 раза ($p < 0.3$). Локальное выравнивание последовательностей между miR-124-3p и выбранными 333 мРНК и 262 днРНК проводили с использованием банка данных NCBI GEO с помощью алгоритма Смита–Уотермана. Были отобраны 8 днРНК и

15 мРНК-мишеней, характеризующихся одним или несколькими сайтами связывания размером 7-8 mer. Определен уровень функционального ответа этих РНК на трансфекцию клеток SKOV3 миметиком miR-124-3p. По данным, полученным на микрочипах Affymetrix HTA 2.0, 5 мРНК-мишеней из 15 предсказанных биоинформатически снизили уровень своей экспрессии в 1.1–1.7 раза (p = 0.01–0.4): EIF2B5, IL15RA, PARP14, RNF169, TRAF3, SNX16. Три днРНК из 8 предсказанных также показали функциональный ответ на сверхэкспрессию miR-124-3p: JAZF1-AS1, TMEM72-AS1, XIST (снижение в 1.02–1.1 раза; p = 0.3–0.8).
Выводы: Идентифицированные днРНК и мРНК, взаимодействующие с miR-124-3p, могут представлять собой новые маркеры РЯ. В дальнейшем потребуется проверка полученных данных на репрезентативной выборке клинических образцов с последующей валидацией на культурах клеток.
Финансирование: Исследование поддержано грантом Российского научного фонда (№ 20-15-00368).

Potential regulatory lncRNAs and mRNA-targets of the critical tumor suppressor miR-124-3p in the ovarian cancer transcriptome

Zaichenko D.M.[1], Astafeva I.R.[1], Uroshlev L.A.[2], Fridman M.V.[2], Burdennyy A.M.[1], Loginov V.I.[1], Moskovtsev A.A.[1]*, Braga E.A.[1]

[1] *Institute of General Pathology and Pathophysiology, Moscow, Russia*

[2] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

* *bioinf@mail.ru*

Key words: transcriptomic data; lncRNA; SKOV3; miR-124-3p overexpression; microarrays Affymetrix HTA 2.0

Motivation and Aim: The motivation for the following work is to identify and characterize the regulatory levels of noncoding RNAs (lncRNAs) and target mRNAs of miR-124-3p in ovarian cancer (OvCa) using transcriptomic data. The goal is to determine the functional response of these RNAs to miR-124-3p overexpression and to identify potential markers of OvCa.
Methods and Algorithms: The work used analysis of transcriptomic data from the NCBI GEO database to screen for lncRNA regulators and miR-124-3p mRNA targets in ovarian cancer, correlation analysis between miR-124-3p and lncRNA/mRNA in the ovarian cancer transcriptome, transfection and overexpression miR-124-3p mimics in SKOV3 cells, analysis of differential expression of lncRNAs and mRNAs in cells using Affymetrix HTA 2.0 microarrays, competitive persistence between miR-124-3p and selected lncRNAs and mRNAs using the Smith–Waterman algorithm, analysis of third-party response of selected lncRNAs and mRNA targets for overexpression of miR-124-3p in SKOV3 cells, bioinformatic prediction of mRNA targets and lncRNAs that interact with miR-124-3p.
Results: The microRNA miR-124-3p has been shown to be a critical tumor-suppressor in many types of cancer. Screening of regulatory lncRNAs and mRNA targets of this miRNA in OvCa was carried out using a transcriptomic data set of 131 patients from NCBI GEO (GSE211669). Sets of 262 lncRNAs (Rs < –0.5, p < 0.01) and 333 mRNAs

(Rs < –0.7, $p < 0.0001$) negatively correlated with miR-124 were identified. Data on correlation analysis RNAs of three types in the OvCa transcriptome were compared with the functional response of SKOV3 cells to overexpression of miR-124-3p. We transfected SKOV3 using a miR-124-3p mimic and analyzed differentially expressed (DE) lncRNAs and mRNAs using Affymetrix HTA 2.0 high-throughput microarrays. Among 333 mRNAs negatively correlated with miR-124, 127 DE mRNAs were detected in SKOV3 cells, of which 94 mRNAs were downregulated by 1.1–1.7-fold ($p < 0.3$). Local sequence alignment between miR-124 and selected 333 mRNAs and 262 lncRNAs was performed using NCBI GEO data bank by the Smith–Waterman algorithm. 8 lncRNAs and 15 mRNA-targets, characterized by one or more binding sites of 7mer-8mer size, were selected. The level of functional response of these RNAs to transfection of SKOV3 cells with a miR-124-3p mimic was determined. According to data obtained on Affymetrix HTA 2.0 microarrays, 5 mRNA-targets out of 15 predicted bioinformatically decreased their expression level by 1.1–1.7 times ($p = 0.01–0.4$): EIF2B5, IL15RA, PARP14, RNF169, TRAF3, SNX16. Three lncRNAs out of 8 predicted also showed a functional response to overexpression of miR-124-3p: JAZF1-AS1, TMEM72-AS1, XIST (1.02–1.1-fold decrease; $p = 0.3–0.8$).

Conclusion: The identified lncRNAs and mRNAs interacting with miR-124-3p may represent new markers of OvCa. In the future, verification of the obtained data on a representative set of clinical samples, followed by validation in cell cultures, is required.

Funding: The study is supported by the Russian Science Foundation grant (No. 20-15-00368).

doi 10.18699/bgrs2024-1.4-06

Анализ данных транскриптомики единичных клеток на примере онкологических заболеваний

Иконников А.В.[1, 4]*, Волчков Е.В.[1, 2], Хозяинова А.А.[3], Меняйло М.Е.[1, 3], Копанцева Е.Е.[1], Гуржиханова М.Х.[2], Порядина Е.А.[1], Фетисов Т.И.[5], Якубовская М.Г.[5], Денисов Е.В.[1, 3], Кирсанов К.И.[5], Масчан М.А.[2]

[1] *НИИ молекулярной и клеточной медицины МИ РУДН, Москва, Россия*

[2] *НМИЦ детской гематологии, онкологии и иммунологии им. Д. Рогачева, Москва, Россия*

[3] *Томский НИМЦ РАН, Томск, Россия*

[4] *Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия*

[5] *ФГБУ «НМИЦ онкологии им. Н.Н. Блохина» Минздрава России, Москва, Россия*

* *alex.v.ikonnikov@gmail.com*

Ключевые слова: биоинформатика; single-cell RNA-seq; рак; ЮММЛ; саркома

Мотивация и цель: Анализ данных секвенирования РНК одиночных клеток (scRNA-seq) позволяет получать информацию о профилях экспрессии опухолевых клеточных популяций и изучать нарушения молекулярных механизмов на уровне отдельных клеток. Понимание клеточной гетерогенности и динамики клеточных взаимодействий в опухоли может привести к разработке новых комбинированных терапевтических стратегий, направленных на более эффективную терапию онкологических заболеваний [1–3].

Целью данного исследования являлся анализ данных РНК-секвенирования единичных клеток образцов пациентов с недифференцированными плеоморфными саркомами и образцов костного мозга (КМ) пациентов с ювенильным миеломоноцитарным лейкозом (ЮММЛ).

Методы и алгоритмы: В качестве образцов группы ЮММЛ были использованы клетки КМ 3 пациентов с мутациями генов PTPN11, NF11, NRAS. Контрольной группой являлись образцы scRNA-секвенирования КМ 3 здоровых детей, полученные из открытой базы данных NCBI GEO. Также было получено 9 образцов клеток тканей пациентов с недифференцированными плеоморфными саркомами. Пробоподготовка для РНК-секвенирования единичных клеток выполнялась на платформе Chromium X (10x Genomics, США). Секвенирование проводилось с помощью Genolab M (GeneMind, Китай). Биоинформатический анализ, включающий фильтрацию клеток и генов, нормализацию данных, интеграцию и типирование клеточных типов, выполнялся с помощью инструментов Scanpy (v.1.9.5) [4] и Seurat (v.5.0.3) [5]. Коррекция батч-эффекта после интеграции образцов было выполнена с помощью пакета Harmony [6]. Анализ метаболических путей выполнялся с использованием пакетов GSEApy [7] и DecoupleR [8].

Результаты: Обнаружено, что образцы ЮММЛ отличались повышенной пролиферацией клеток миеломоноцитарного ряда в сравнении с контролем. Образец ЮММЛ с агрессивным клиническим течением и мутацией в PTPN11 отличался значительным количеством стволовых клеток (HSCs), ранних предшественников и клеток эритроидного ряда. Анализ метаболических путей выявил, что клетки образцов ЮММЛ характеризовались повышенной активностью метаболического пути MAPK в сравнении с контролем. Данная

активность была неравномерна в клеточных популяциях и наиболее выражена в HSCs и MEPs (megakaryocyteerythroid progenitor cells). Также была обнаружена гиперактивация PI3K-пути в образце ЮММЛ с мутацией *PTPN11*. Выявлено преобладание активности путей, связанных с ГТФазами и деацетилированием гистонов в клетках GMPs (granulocyte/monocyte progenitors), в образцах ЮММЛ в сравнении с контролем.
С помощью биоинформатической аннотации был охарактеризован клеточный ландшафт образцов ткани недифференцированных плеоморфных сарком. Обнаружены кластеры опухоль-ассоциированных фибробластов (CARs), опухолевых клеток, макрофагов, Т клеток и минорных редких клеточных популяций, характерные для отдельных пациентов.
Выводы: В результате исследования были изучены профили экспрессии генов различных клеточных популяций образцов КМ ЮММЛ в сравнении с образцами КМ контрольной группы. Были идентифицированы клеточные типы, присутствующие в образцах недифференцированных плеоморфных сарком.
Финансирование: Работа выполнена при финансовой поддержке Российского научного фонда (проект № 23-65-00003).

Analysis of single-cell transcriptomics data using the example of cancer diseases

Ikonnikov A.V.[1, 4]*, Volchkov E.V.[1, 2], Khozyainova A.A.[3], Menyailo M.E.[1, 3], Kopantseva E.E.[1], Gurzhikhanova M.Kh.[2], Poryadina E.A.[1], Fetisov T.I.[5], Yakubovskaya M.G.[5], Denisov E.V.[1, 3], Kirsanov K.I.[5], Maschan M.A.[2]

[1] *Research Institute of Molecular and Cellular Medicine, Peoples' Friendship University of Russia (RUDN University), Moscow, Russia*

[2] *Dmitry Rogachev National Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia*

[3] *Tomsk National Research Medical Center (TNRMC), Tomsk, Russia*

[4] *Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia*

[5] *Federal State Budgetary Institution, "N.N. Blokhin National Medical Research Center of Oncology" of the Ministry of Health of the Russian Federation (N.N. Blokhin NMRCO), Moscow, Russia*

* *alex.v.ikonnikov@gmail.com*

Key words: bioinformatics; single-cell RNA-seq; cancer; JMML; sarcoma

Motivation and Aim: Single-cell RNA sequencing (scRNA-seq) analysis allows for the characterization of expression profiles of tumor cell populations and the study of disruptions in molecular mechanisms at the level of individual cells. Understanding cellular heterogeneity and dynamics of cell-cell interactions in tumors may lead to the development of novel combined therapeutic strategies aimed at more effective treatment of oncological diseases [1–3].
The aim of this study was to analyze single-cell RNA-seq data from samples obtained from patients with undifferentiated pleomorphic sarcomas and bone marrow (BM) samples obtained from patients with juvenile myelomonocytic leukaemia (JMML).
Methods and Algorithms: Bone marrow cells from 3 patients with mutations in the PTPN11, NF11 and NRAS genes were used as samples from the JMML group. The

control group consisted of bone marrow samples obtained by scRNA sequencing from 3 healthy children obtained from the NCBI GEO open database. 9 tissue cell samples were also obtained from patients with undifferentiated pleomorphic sarcomas. The preparation of samples for single-cell RNA analysis was carried out on the Chromium X platform (10x Genomics, USA). Sequencing was performed using Geolab M (GeneMind, China). Bioinformatic analysis, including cell and gene filtering, data normalization, integration and cell type annotation, was performed using the Scanpy (version 1.9.5) [4] and Seurat (version 5.0.3) [5] packages. Correction of the batch effect after sample integration was performed using the Harmony [6] package. The analysis of metabolic pathways was conducted using the GSEApy [7] and DecoupleR [8] packages.
Results: It was found that JMML samples were characterized by increased proliferation of myelomonocytic cells compared to the control. A JMML sample with an aggressive clinical course and a mutation in PTPN11 was characterized by significant numbers of stem cells (HSCs), early progenitors, and erythroid cells. Analysis of metabolic pathways revealed that cells from JMML samples were characterized by increased activity of the MAPK metabolic pathway compared to controls. This activity was uneven across cell populations and was most pronounced in HSCs and MEPs (megakaryocyteerythroid progenitor cells). Hyperactivation of the PI3K pathway was also detected in a JMML sample with a PTPN11 mutation. The predominance of the activity of pathways associated with GTPases and histone deacetylation in GMPs (granulocyte/monocyte progenitors) cells in JMML samples was revealed in comparison with the control.
As a result of bioinformatics annotation, the cellular landscape of undifferentiated pleomorphic sarcoma samples was characterized. Clusters of tumor-associated fibroblasts (CARs), tumor cells, macrophages, T cells and minor rare cell population characteristic of individual patients were detected.
Conclusion: The study analyzed the gene expression profiles of various cell populations from JMML BM samples, comparing them with healthy control BM samples. The cellular types present in samples of undifferentiated pleomorphic sarcomas were identified.
Funding: The study is supported by the Russian Science Foundation (project No. 23-65-00003).

Список литературы/References

1. Chen S. et al. Single-cell analysis technologies for cancer research: from tumor-specific single cell discovery to cancer therapy. *Front Genet.* 2023;14:1276959
2. Louka E. et al. Heterogeneous disease-propagating stem cells in juvenile myelomonocytic leukemia. *J Exp Med.* 2021;218(2):e20180853
3. Yuan L.L. et al. Single-cell sequencing reveals the landscape of the tumor microenvironment in a skeletal undifferentiated pleomorphic sarcoma patient. *Front Immunol.* 2022;13:1019870. doi 10.3389/fimmu.2022.1019870
4. Wolf F.A. et al. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*. 2018;19(1):15. doi 10.1186/s13059-017-1382-0
5. Hao Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. 2024;42(2):293-304
6. Korsunsky I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16(12):1289-1296
7. Fang Z. et al. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics*. 2023;39(1):btac757
8. Badia-I-Mompel P. et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv*. 2022;2(1):vbac016. doi 10.1093/bioadv/vbac016

doi 10.18699/bgrs2024-1.4-07

Транскриптомный анализ развития полипида пресноводной мшанки *Cristatella mucedo*: от почкования до дегенерации

Квач А.Ю.[1]*, Кутюмов В.А.[1], Старунов В.В.[1, 2], Островский А.Н.[1]

[1] *Санкт-Петербургский государственный университет, кафедра зоологии беспозвоночных, Санкт-Петербург, Россия*

[2] *Зоологический институт РАН, Лаборатория эволюционной морфологии, Санкт-Петербург, Россия*

* *ay.kvach@yandex.ru*

Ключевые слова: RNA-seq; развитие; колониальность; Bryozoa; почкование, дифференциальная экспрессия генов

Мотивация и цель: Изучение колониальных беспозвоночных позволяет взглянуть на ряд важных биологических процессов под новым углом. Для роста и развития модульного организма необходимы: (1) механизмы, обеспечивающие почкование модулей на протяжении всей жизни колонии за счет наличия стволовых клеток или клеточной дедифференцировки; (2) механизмы периодической дегенерации модулей колонии, сроки жизни которых существенно короче, чем у самой колонии; (3) механизмы, регулирующие эти антагонистические процессы. Колониальная организация возникала неоднократно и независимо в пяти типах Eumetazoa (Cnidaria, Bryozoa, Kamptozoa, Hemichordata, Chordata) [1], однако за исключением нескольких модельных видов гидроидов и асцидий, нам очень мало известно о развитии колониальных беспозвоночных на молекулярном и клеточном уровнях. С целью восполнить этот пробел для Bryozoa, мы провели постадийный транскриптомный анализ развития полипида в колонии пресноводной мшанки *Cristatella mucedo*.

Методы и алгоритмы: Мы секвенировали транскриптомы 6 последовательных стадий развития и деградации полипида – пищедобывающего щупальцевого аппарата с кишечником, включая почку полипида. ювенильный полипид, три возраста зрелых полипидов и дегенерирующий полипид, на платформе Illumina HiSeq2500. На основе полученных данных мы собрали и проаннотировали референсный транскриптом. После этого мы провели анализ дифференциальной экспрессии генов, выявили кластеры ко-экспрессии и охарактеризовали их с помощью анализов обогащения GO и KO групп. Также мы исследовали паттерны экспрессии наиболее важных групп генов, связанных с процессами развития и их регуляцией.

Результаты: Подавляющие большинство консервативных маркеров стволовых клеток и транскрипционных факторов, связанных с процессами развития, а также элементы Wnt/beta-catenin сигналинга, повышают свою экспрессию на начальных стадиях формирования полипида. Однако небольшая доля из них также повышают экспрессию в зрелых или дегенерирующих полипидах. Наши результаты свидетельствуют о том, что дегенерация полипидов – сложный регулируемый процесс, включающий аутофагию, апоптоз, и другие типы программируемой

клеточной смерти, а также, вероятно, локальную транс- или дедифференцировку клеток. Также мы выдвигаем гипотезу, что сигнальный путь mTOR может играть важную роль в регулировании продолжительности жизни полипидов, а также переключении между фазами «развития», «функционирования» и «дегенерации».
Выводы: Проанализировав изменения в транскриптоме полипида во время его развития, мы выявили круг генов, процессов и сигнальных путей, обеспечивающих как формирование, так и дегенерацию полипида. Кроме того, мы обнаружили, что дегенерация полипида, по всей видимости, такой же сложный и регулируемый процесс, как и почкование, и, таким образом, представляет собой перспективную область для дальнейших исследований.
Финансирование: Исследование поддержано грантом РНФ № 23-14-00351.

Transcriptome analysis of polypide development in the freshwater bryozoan *Cristatella mucedo*: from budding to degeneration

Kvach A.[1]*, Kutyumov V.[1], Starunov V.[1, 2], Ostrovsky A.[1]

[1] *Saint Petersburg State University, Department of Invertebrate Zoology, Saint-Petersburg, Russia*

[2] *Zoological Institute of RAS, Laboratory of Evolutionary Morphology, Saint-Petersburg, Russia*

** ay.kvach@yandex.ru*

Key words: RNA-seq; development; coloniality; Bryozoa; budding, differential gene expression

Motivation and Aim: The study of colonial invertebrates provides a new perspective on a number of important biological processes. Growth and maintenance of the colonial body is provided by (1) mechanisms that ensure budding of modules throughout the life of the colony using stem cell lineage or the ability of the cells to dedifferentiate; (2) mechanisms for the cyclic non-pathological degeneration of colony modules whose lifespan is shorter than that of the colony itself; (3) mechanisms that regulate these antagonistic processes. Colonial body organization has arisen repeatedly and independently in five eumetazoan phyla (Cnidaria, Bryozoa, Kamptozoa, Hemichordata, Chordata) [1], but with the exception of a few model species of hydroids and ascidians, we know very little about the basics of coloniality at the molecular and cellular levels. Our study on the freshwater *Cristatella mucedo* is the first step in addressing this gap to Bryozoa.
Methods and Algorithms: We sequenced the transcriptomes of 6 consecutive stages of the development of the polypide – tentacle food-gathering apparatus with a gut), namely the polypide bud, juvenile polypide, three ages of mature polypides, and degenerating polypide, using the Illumina HiSeq2500 platform. Next, we assembled and annotated the reference de novo transcriptome. After that we performed differential gene expression analysis, identified gene co-expression clusters and characterized them using GO and KO group enrichment analyses. We also investigated the expression patterns of the most important groups of genes associated with developmental processes and their regulation.
Results: The vast majority of conserved "stemness" markers and transcription factors, associated with developmental processes, as well as elements of the Wnt/beta-catenin signaling pathway increase their expression during the early stages of polypide

development. However, a small amount of them also increase expression in mature or degenerating polypides. Our results suggest that polypide degeneration is a complex regulated process involving autophagy, apoptosis, and other types of programmed cell death, as well as probably local trans- or dedifferentiation of cells. We also suggest that the mTOR signaling pathway may play an important role in regulating the lifespan of polypides.

Conclusion: By analyzing changes in the polypide transcriptome during development, we have identified the genes, processes and signaling pathways that enable both budding and polypide degeneration. In addition, we found that polypide degeneration appears to be as complex and regulated as budding and, thus, represents a promising area for further research.

Funding: The study has been supported by the Russian Science Foundation (grant No. 23-14-00351).

Список литературы/References

1. Hiebert L.S., Simpson C., Tiozzo S. Coloniality, clonality, and modularity in animals: The elephant in the room. *J Exp Zool Mol Dev Evol.* 2021;336(3):198-211

doi 10.18699/bgrs2024-1.4-08

Определение генетической сигнатуры химически поляризованных макрофагов крысы

Киселева В.[1, 2]*, Вишнякова П.[1, 2], Лохонина А.[1, 2, 3], Машкова О.[4], Карпулевич Е.[4], Фатхудинов Т.[1, 2]

[1] *ФГБУ «Национальный медицинский исследовательский центр акушерства, гинекологии и перинатологии им. академика В.И. Кулакова» Министерства здравоохранения Российской Федерации, Москва, Россия*

[2] *ФГАОУ ВО «Российский университет дружбы народов им. Патриса Лумумбы», Москва, Россия*

[3] *НИИ морфологии человека им. академика А.П. Авцына ФГБНУ «Российский научный центр хирургии им. академика Б.В. Петровского», Москва, Россия*

[4] *Федеральное государственное бюджетное учреждение науки Институт системного программирования им. В.П. Иванникова РАН, Москва, Россия*

* *Victoria.kurnosova.1991@gmail.com*

Ключевые слова: макрофаги; дифференциально экспрессируемые гены; M(ЛПС); M(IL4/IL10)

Мотивация и цель: Макрофаги – клетки врожденного иммунного ответа, которые в зависимости от окружающих условий могут проявлять два противоположных состояния – провоспалительное и противовоспалительное. В эру персонализированной медицины и развития клеточных технологий они представляют большой интерес, так как играют значительную роль в патогенезе многих заболеваний. Изучение роли и потенциального терапевтического эффекта макрофагов часто проводят на модельных животных, являющихся мостиком между наукой и клиникой. И если иммунные процессы мыши описаны довольно подробно, то для крысы остается много белых пятен. Так, в литературе встречаются единичные работы, посвященные изучению механизмов поляризации макрофагов, происходящих из костного мозга, на транскрипционном уровне [1]. Поэтому целью данной работы стало определение генетических сигнатур макрофагов крысы, поляризованных *in vitro* провоспалительными и противовоспалительными стимулами.

Методы и алгоритмы: Моноциты крысы выделяли из костного мозга при градиентном центрифугировании на препарате синтетического высокомолекулярного сополимера полисахарозы (Фиколл) с последующем иммуномагнитным разделением по наличию лейкосиалина (CD43) (трансмембранного гликопротеина, играющего важную роль в клеточной адгезии, дифференцировке и апоптозе). Полученные моноциты дифференцировали в макрофаги в течение 7 дней в присутствии MCSF и далее подвергали химической модификации с помощью ЛПС (провоспалительный ответ, M(ЛПС)) и IL-4/IL-10 (противовоспалительный ответ, M(IL-4/IL-10)) в течение 24 часов. После поляризации клетки были лизированы для дальнейшего анализа транскриптома методом секвенирования следующего поколения (NGS). «Сырые» NGS данные были обработаны несколькими биоинформатическими алгоритмами (*DeSeq2*, *voom+limma*, *EBSeq*, *NOISeq*, *edgeR)*. Количественные результаты по генам, полученные применением инструмента *Salmon* [2] к множественному выравниванию, выполненному с помощью STAR [3], были использованы в

пайплайне Hobotnica [4] для оценки выбора лучшего инструмента для анализа дифференциальной экспрессии генов. В соответствии с критерием выбора лучшего инструмента для анализа дифференциальной экспрессии генов [4] был выбран *DeSeq2*, результаты которого были отфильтрованы по порогу 0.1 для скорректированных p-значений. Для демонстрации общего результата использовался volcano plot сравнения дифференциальной экспрессии групп M0 и M(ЛПС), M0 и M(IL-4/IL-10)) соответственно.
Результаты: Выявлены значительные изменения в экспрессии более сотни генов. Анализ обогащения (enrichment анализ) выявил, что как после стимуляции ЛПС, так и после IL-4/IL-10 поляризации наиболее представленным сигнальным путем был каскад, ассоциированный с конверсией метаболитов (metabolite interconversion enzyme). Следующими по представленности были: белок-модифицирующий каскад (protein modifying enzyme) и сигнальный путь, ассоциированный с трансмембранными сигнальными рецепторами (transmembrane signal receptor) – среди значимо повышенных и пониженных генов соответственно и в M(ЛПС) макрофагах по сравнению с M0 макрофагами. В случае значимо повышенных и пониженных генов в M(IL-4/IL-10) макрофагах по сравнению с M0 макрофагами помимо упомянутых путей были обнаружены каскады, ассоциированные с ген-специфическими регуляторами транскрипции (gene-specific transcriptional regulator) и транспортерами (transporter).
Выводы: В результате проделанной работы была определена генетическая сигнатура макрофагов с провоспалительным и противовоспалительным фенотипом. Были выявлены наиболее активированные сигнальные пути при стимуляции ЛПС и IL-4/IL-10. Результаты секвенирования доступны в GEO (GSE266262).
Финансирование: Исследование поддержано Российским научным фондом (номер гранта 24-25-00203).

Determination of the genetic signature of chemically polarized rat macrophages

Kiseleva V.[1, 2]*, Vishnyakova P.[1, 2], Lokhonina A.[1, 2, 3], Mashkova O.[4], Karpulevich E.[4], Fatkhudinov T.[1, 2]

[1] *Federal State Budget Institution "National Medical Research Center for Obstetrics, Gynecology and Perinatology named after Academician V.I. Kulakov", Ministry of Healthcare of the Russian Federation, Moscow, Russia*

[2] *Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University), Moscow, Russia*

[3] *Avtsyn Research Institute of Human Morphology of Federal State Budgetary Scientific Institution "Petrovsky National Research Centre of Surgery", Moscow, Russia*

[4] *Federal State Budgetary Institution of Science V.P. Ivannikov Institute of System Programming, RAS, Moscow, Russia*

** Victoria.kurnosova.1991@gmail.com*

Key words: macrophages; differentially expressed genes; M(LPS); M(IL4/IL10)

Motivation and Aim: Macrophages, the cells of the innate immune response, can exhibit two opposite states, pro-inflammatory and anti-inflammatory, depending on the surrounding conditions. In the era of personalised medicine and the development of cellular technologies, they are of great interest as they play a significant role in the pathogenesis of many diseases. Studies of the role and potential therapeutic effects of macrophages are often carried out on animal models, which are the bridge between science and the clinic. And if the immune processes of the mouse are described in quite detail, there are still many white spots for the rat. Thus, in the literature there are single works devoted to the study of mechanisms of polarization of bone marrow-derived macrophages at the transcriptional level [1]. Therefore, the aim of this work was to determine the genetic signatures of rat macrophages polarized *in vitro* by pro-inflammatory and anti-inflammatory stimuli.
Methods and Algorithms: Rat monocytes were isolated from bone marrow by gradient centrifugation on a preparation of synthetic high-molecular-weight polysaccharose copolymer (Ficoll) followed by immunomagnetic separation for the presence of leukosialin (CD43) (a transmembrane glycoprotein that plays an important role in cell adhesion, differentiation and apoptosis). The resulting monocytes were differentiated into macrophages for 7 days in the presence of MCSF and then chemically modified with LPS (pro-inflammatory response, M(LPS)) and IL-4/IL-10 (anti-inflammatory response, M(IL-4/IL-10)) for 24 hours. After polarization, cells were lysed for further analysis of the transcriptome by next generation sequencing (NGS). 'Raw' NGS data were processed by several bioinformatics algorithms (DeSeq2, voom + limma, EBSeq, NOISeq, edgeR). The quantitative gene results obtained by applying the Salmon tool [2] to the multiple alignment performed with STAR [3] were used in the Hobotnica Pipeline [4] to evaluate the choice of the best tool for differential gene expression analysis. According to the criterion for selecting the best tool for analysing differential gene expression [4], DeSeq2 was selected and the results were filtered by a threshold of 0.1 for adjusted p-values. The volcano plot comparing differential expression of M0 and M(LPS), M0 and M(IL-4/IL-10) groups respectively was used to demonstrate the overall result.
Results: Significant changes in the expression of more than a hundred genes were detected. Enrichment analysis revealed that both after LPS stimulation and IL-4/IL-10 polarisation the most represented signal pathway was the cascade associated with metabolite interconversion (metabolite interconversion enzyme). The protein modifying enzyme and the sig- nal pathway associated with transmembrane signal receptors were the next most represented among the significantly up- and down-regulated genes, respectively, and in M(LPS) macrophages compared to M0 macrophages. In the case of significantly upregulated and downregulated genes in M(IL-4/IL-10) macrophages compared to M0 macrophages, cascades associated with gene-specific transcriptional regulators (gene-specific transcriptional regulator) and transporters (transporter) were detected in addition to the above-mentioned pathways.
Conclusion: The genetic signature of macrophages with proinflammatory and anti-inflammatory phenotype was determined. The most activated signalling pathways under LPS and IL-4/IL-10 stimulation were identified. Sequencing results are available at GEO (GSE266262).
Funding: This work was supported by Russian Science Foundation (grant number 24-25-00203).

Список литературы/References

1. Pridans C., Irvine K.M., Davis G.M., Lefevre L., Bush S.J., Hume D.A. Transcriptomic analysis of rat macrophages. *Front Immunol.* 2021;11:594594. doi 10.3389/fimmu.2020.594594
2. Patro R., Duggal G., Love M. et al. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017;14:417-419. doi 10.1038/nmeth.4197
3. Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi 10.1093/bioinformatics/bts635
4. Stupnikov A., McInerney C.E., Savage K.I., McIntosh S.A., Emmert-Streib F., Kennedy R., Salto-Tellez M., Prise K.M., McArt D.G. Robustness of differential gene expression analysis of RNA-seq. *Comput Struct Biotechnol J.* 2021;19:3470-3481. doi 10.1016/j.csbj.2021.05.040

doi 10.18699/bgrs2024-1.4-09

Транскриптомный ответ *Lacticaseibacillus rhamnosus* на цитокины IL-8 и IL-10

Климина К.[1, 2]*, Дьячкова М.[2], Веселовский В.[1], Захаревич Н.[1], Олехнович Е.[1]

[1] *Федеральный научно-клинический центр Физико-химической медицины им. академика Ю.М. Лопухина Федерального медико-биологического агентства, Москва, Россия*

[2] *Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия*

* *ppp843@yandex.ru*

Ключевые слова: комменсальная микробиота кишечника; иммуномодулирующие бактерии; иммунный ответ хозяина, цитокины; РНК-секвенирование; транскриптом; филогенетическое профилирование; фосфотрансферазная система

Мотивация и цель: Современные исследования подчеркивают динамическое взаимодействие между микробиотой кишечника и иммунной системой. Иммунные факторы, включая цитокины, оказывают значительное влияние на разнообразие кишечной микробиоты, влияя на ее состав и функциональность, что, в свою очередь, влияет на здоровье и иммунитет. Особое внимание стоит уделить лактобациллам и бифидобактериям, обладающим выраженными иммуномодулирующими свойствами. Эти микроорганизмы адаптируются к воздействию цитокинов, что способствует их выживанию и функционированию в кишечной среде. Данное исследование направлено на понимание механизмов взаимодействия микробиоты и организма хозяина, рассматривая микроорганизмы не только как возможных индикаторов изменений в состоянии хозяина, но и как активных участников в иммунном ответе. Это открывает новые подходы к регуляции иммунной системы через манипуляцию микробиоты и предлагает новые стратегии для улучшения здоровья человека. Однако ключевым аспектом работы является изучение двусторонней системы передачи сигналов, развившихся в процессе коэволюции организма хозяина и его микробиоты. В частности, исследуются короткие пептидные последовательности – цитокины, которые могут сигнализировать о воспалительных процессах и позволять микробиоте адаптироваться к предстоящему иммунному ответу, тем самым способствуя ее выживанию.

Методы и алгоритмы: В данной работе использовали штамм *L. rhamnosus* K32 (GenBank JNNV000000). Культивирование проводили в бульоне Lactobacillus MRS Broth (HiMedia) при 37 °C в анаэробных условиях (HiAnaerobic System, HiMedia). Ночную культуру штамма разводили в соотношении 1:100 в 2 мл среды. Экспериментальные образцы обрабатывали лиофилизированными рекомбинантными человеческими цитокинами – IL-8 и IL-10 в концентрациях 1 нг/мл (Thermo Fisher Scientific). Контрольная группа состояла из штамма *L. rhamnosus*, выращенного на среде MRS с добавлением деионизированной воды и рекомбинантного белка GFP (Thermo Fisher Scientific). Транскриптомный анализ проводили в экспоненциальной фазе роста (OD600 = 1.2). Эксперимент проводили в трех повторах. Тотальную РНК выделяли из 2 мл культуры клеток, предварительно обработав 1 мл реагента RNAprotect Bacteria Reagent (Qiagen) для стабилизации РНК. Клетки разрушали в 2 мл пробирках Lysing Matrix B с помощью прибора MagNA Lyzer (Roche) в течение 30 секунд. Для выделения РНК

использовали автоматическую станцию KingFisher Flex и набор MagMAX mirVana Total RNA Isolation Kit (ThermoFisher Scientific). Количество и качество выделенной РНК оценивали на флуориметре Qubit с набором Quant-it RiboGreen RNA assay kit (Thermo Fisher Scientific) и Bioanalyzer 2100 с чипом RNA 6000 Pico (Agilent Technologies) соответственно. Для приготовления транскриптомных библиотек брали 200 нг тотальной РНК. Удаление рибосомальной РНК проводили с помощью набора Ribo-Zero Plus rRNA Depletion Kit (Illumina), библиотеки готовили с помощью набора NEBNext Ultra II Directional RNA Library Prep Kit (NEB). Качество и размер библиотек оценивали с помощью набора High Sensitivity DNA Kit (Agilent Technologies), а концентрацию библиотек измеряли набором Quant-iT DNA Assay Kit, High Sensitivity (Thermo Fisher Scientific). После этого эквимолярно объединяли библиотеки и разводили до конечной концентрации 12 пМ. Секвенирование проводили на платформе HiSeq 2500 (Illumina) с использованием набора HiSeq Rapid SBS Kit v2 (50 циклов) и HiSeq SR Rapid Cluster Kit v2, добавляли 1 % Phix (Illumina) в качестве внутреннего контроля. Контроль качества прочтений проводили с помощью FastQC v0.12.1. Выравнивание на референсный геном – с помощью HISAT2 v2.2.1. Оценку качества сопоставленных прочтений проводили QualiMap v.2.2.2-dev. Конвертацию из формата SAM в формат BAM, а также их сортировку и индексацию делали с помощью SAMtools v1.18. Подсчет прочтений, согласованных с геномными аннотациями, – с помощью featureCounts v2.0.6. Дифференциальную экспрессию генов (ДЭГ) определяли с помощью пакета edgeR 3.40.2 для R v4.1.2. Функциональная аннотация ДЭГ с помощью кластера ортологичных генов (Cluster of Orthologous Genes, COG) с использованием инструмента reCOGnizer v1.10.1. Филогенетическое профилирование состояло из нескольких этапов: сначала определяли ортогруппы в геномах *Lacticaseibacillus*. Далее строились бинарные векторы для обозначения присутствия или отсутствия в геномах. Затем была создана матрица парных расстояний между этими векторами, после чего была проведена группировка филогенетических профилей (ФП) в программе OrthoFinder v.2.5.4. ФП были созданы для 163 полных геномов представителей *Lacticaseibacillus*, включая *L. rhamnosus* K32. Попарное расстояние между ФП рассчитывали с помощью меры Жаккара. Гены объединяли в группы на основании минимальных парных расстояний между ФП (отсечение = 0.001) в матрице расстояний, физической близости (в пределах 10 000 п. н.) в пределах одной ветви и совместной регуляции всех генов в группах. Для этого анализа был использован собственный скрипт Python v.3.10.2 (github.com/LabGenMO/phylo-profiling/blob/master/Identification%20of%20connected%20genes.ipynb).

Результаты: В ходе исследования проведена оценка воздействия цитокинов IL-8 и IL-10 на культуру *Lacticaseibacillus rhamnosus*. Наши эксперименты показали, что присутствие в среде цитокинов IL-8 и IL-10 не влияет на скорость роста культуры. Однако эти цитокины могут влиять на экспрессию генов, не связанных с ростом клеток. Для дальнейшего понимания механизмов, лежащих в основе реакции лактобацилл на воспаление, мы провели высокопроизводительное секвенирование транскриптома штамма *L. rhamnosus* K32 в экспоненциальной фазе роста в экспериментальных и контрольных условиях в трех независимых биологических повторах. Было выявлено 57 ДЭГ для IL-8 и 72 для IL-10, экспрессия которых изменялась более чем в 2 раза ($FC \geq |2|$, $FDR < 0.05$). Функциональный анализ ДЭГ с пониженной экспрессией показал изменения в

категориях K (транскрипция), G (транспорт и метаболизм углеводов) и C (производство и преобразование энергии). В этих группах наблюдалось большое количество ДЭГ, общих для IL-8 и IL-10. Кроме того, для IL-10 наблюдалось повышение экспрессии гена (QJQ50_RS00105), относящегося к категории V (защитные механизмы), а для IL-8 – гена (QJQ50_RS04340), относящегося к категории M (Биогенез клеточной стенки/мембраны/оболочки). Анализ выявил значительную долю генов, связанных с фосфотрансферазной системой, которые демонстрировали снижение экспрессии после обработки обоими цитокинами. Эта система хорошо изучена с точки зрения ее участия в поглощении и утилизации углеводных субстратов. Наблюдаемое снижение экспрессии генов, связанных с транспортом и метаболизмом углеводов, производством и преобразованием энергии, а также транскрипцией, отражает уменьшение метаболической активности *L. rhamnosus* K32. Это указывает на адаптацию культуры в ответ на воздействие цитокинов IL-8 и IL-10, что является защитной реакцией на измененные условия окружающей среды. Для выявления предполагаемых оперонов в ДЭГ мы использовали ФП. Было выявлено десять предполагаемых оперонов в образцах, обработанных IL-8, и девять оперонов в образцах, обработанных IL-10, содержащих два или более функционально связанных гена. Обнаружение групп генов в одном опероне под воздействием IL-8 и IL-10 предполагает, что в ответ на эти цитокины бактерии могут регулировать ключевые биохимические пути. Это может иметь значительные последствия для их метаболизма, ответа на стресс и взаимодействия с иммунной системой хозяина. В свою очередь, это может способствовать адаптации бактерий к воспалительной среде или модуляции воспалительного процесса.
Выводы: Полученные результаты показывают, как воздействие противовоспалительных и провоспалительных цитокинов изменяет профили экспрессии у штаммов *L. rhamnosus* K32. Снижение активности генов, задействованных в энергетических процессах, таких как углеводный обмен и транспорт, стимулирует переход клеток в режим минимального взаимодействия с внешней средой. Этот адаптивный переход способствует экономии и перераспределению ресурсов для поддержания жизнеспособности клеток и активирует защитные механизмы в условиях воспалительного стресса.
Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 22-75-10029, https://rscf.ru/project/22-75-10029/.

Transcriptional responses of *Lacticaseibacillus rhamnosus* to IL-8 and IL-10 cytokines

Klimina K.[1, 2]*, Dyachkova M.[2], Veselovsky V.[1], Zakharevich N.[1], Olekhnovich E.[1]

[1] *Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine, Moscow, Russia*

[2] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

* *ppp843@yandex.ru*

Key words: commensal gut microbiota; immunomodulatory bacteria; host immune response; cytokine signaling; RNA-sequencing; transcriptome; phylogenetic profiling; phosphotransferase system

Motivation and Aim: Current research highlights the dynamic interaction between the gut microbiota and the immune system. Immune factors, including cytokines,

significantly influence the diversity of the gut microbiota, affecting its composition and functionality, which in turn impacts health and immunity. Special attention should be given to lactobacilli and bifidobacteria, which possess pronounced immunomodulatory properties. These microorganisms adapt to the effects of cytokines, which aids their survival and function in the gut environment. This study aims to understand the mechanisms of interaction between the microbiota and the host organism, considering microorganisms not only as potential indicators of changes in the host's condition but also as active participants in the immune response. This opens up new approaches to regulating the immune system through manipulation of the microbiota and offers new strategies for improving human health. However, a key aspect of this research is the study of the bidirectional signaling system that has evolved through the coevolution of the host and its microbiota. In particular, short peptide sequences (cytokines) are investigated, which can signal inflammatory processes and allow the microbiota to adapt to an upcoming immune response, thereby facilitating its survival.

Methods and Algorithms: In this study, we used the strain *L. rhamnosus* K32 (GenBank JNNV000000). Cultivation was carried out in Lactobacillus MRS Broth (HiMedia) at 37 °C under anaerobic conditions (HiAnaerobic System, HiMedia). The overnight culture of the strain was diluted in a ratio of 1:100 in 2 ml of medium. Experimental samples were treated with lyophilized recombinant human cytokines – IL-8 and IL-10 at concentrations of 1 ng/ml (Thermo Fisher Scientific). The control group consisted of *L. rhamnosus* cultured in MRS medium with the addition of deionized water and recombinant GFP protein (Thermo Fisher Scientific). Transcriptomic analysis was conducted during the exponential growth phase (OD600 = 1.2). The experiment was carried out in triplicate. Total RNA was extracted from 2 ml of cell culture, after treating 1 ml with RNAprotect Bacteria Reagent (Qiagen) for RNA stabilization. Cells were lysed in 2 ml Lysing Matrix B tubes using the MagNA Lyzer (Roche) for 30 seconds. RNA was extracted using the KingFisher Flex automated station and the MagMAX mirVana Total RNA Isolation Kit (ThermoFisher Scientific). The quantity and quality of the extracted RNA were assessed on a Qubit fluorometer with the Quant-it RiboGreen RNA assay kit (Thermo Fisher Scientific) and the Bioanalyzer 2100 with the RNA 6000 Pico chip (Agilent Technologies), respectively. For the preparation of transcriptomic libraries, 200 ng of total RNA was used. Ribosomal RNA removal was performed using the Ribo-Zero Plus rRNA Depletion Kit (Illumina), and libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit (NEB). The quality and size of the libraries were evaluated using the High Sensitivity DNA Kit (Agilent Technologies), and the library concentration was measured with the Quant-iT DNA Assay Kit, High Sensitivity (Thermo Fisher Scientific). Libraries were then equimolarly pooled and diluted to a final concentration of 12 pM. Sequencing was performed on the HiSeq 2500 platform (Illumina) using the HiSeq Rapid SBS Kit v2 (50 cycles) and HiSeq SR Rapid Cluster Kit v2, with 1 % Phix (Illumina) added as an internal control. Quality control of the reads was conducted using FastQC v0.12.1. Alignment to the reference genome was performed using HISAT2 v2.2.1. Quality assessment of aligned reads was done with QualiMap v.2.2.2-dev. Conversion from SAM to BAM format, as well as their sorting and indexing, were performed using SAMtools v1.18. Counting of reads aligned with genomic annotations was done using featureCounts v2.0.6. Differential gene expression (DGE) was determined using the edgeR package 3.40.2 for R v4.1.2. Functional annotation of DGE was performed using the cluster of orthologous genes (COG) system with the tool reCOGnizer v1.10.1. Phylogenetic profiling consisted of several stages:

firstly, ortho groups were identified in Lacticaseibacillus genomes. Binary vectors were then constructed to denote the presence or absence in genomes. A matrix of pairwise distances between these vectors was created, after which phylogenetic profiles (PP) were clustered using OrthoFinder v.2.5.4. PPs were created for 163 complete genomes of *Lacticaseibacillus* representatives, including *L. rhamnosus* K32. Pairwise distances between PPs were calculated using the Jaccard measure. Genes were grouped based on the minimal pairwise distances in the distance matrix (cutoff = 0.001), physical proximity (within 10,000 bp) within a branch, and co-regulation of all genes in groups. This analysis was performed using a custom Python script v.3.10.2 (github.com/LabGenMO/ phylo-profiling/blob/master/Identification%20of%20connected%20genes.ipynb).

Results: During the study, we evaluated the impact of cytokines IL-8 and IL-10 on the culture of *Lacticaseibacillus rhamnosus*. Our experiments demonstrated that the presence of IL-8 and IL-10 in the medium does not affect the growth rate of the culture. However, these cytokines can influence the expression of genes unrelated to cell growth. To further understand the mechanisms underlying the lactobacilli's response to inflammation, we performed transcriptomic sequencing of the *L. rhamnosus* K32 strain during its exponential growth phase under experimental and control conditions in three independent biological replicates. We identified 57 differentially expressed genes (DEGs) for IL-8 and 72 for IL-10, with expression changes of more than two-fold ($FC \geq |2|$, $FDR < 0.05$). Functional analysis of the downregulated DEGs revealed changes in categories K (transcription), G (carbohydrate transport and metabolism), and C (energy production and conversion). These groups showed a significant number of DEGs common to both IL-8 and IL-10. Additionally, for IL-10, there was an increase in the expression of gene (QJQ50_RS00105) belonging to category V (defensive mechanisms), and for IL-8, an increase in gene (QJQ50_RS04340) belonging to category M (biogenesis of cell wall/membrane/envelope). The analysis revealed a significant proportion of genes associated with the phosphotransferase system, which downregulated after treatment with both cytokines. This system is well-studied for its role in the uptake and utilization of carbohydrate substrates. The observed decrease in the expression of genes related to carbohydrate transport and metabolism, energy production and conversion, and transcription reflects a decrease in the metabolic activity of *L. rhamnosus* K32. This indicates an adaptation of the culture in response to the cytokines IL-8 and IL-10, which is a protective response to the changed environmental conditions. To identify putative operons in DEGs, we used phylogenetic profiling. Ten putative operons were identified in samples treated with IL-8, and nine operons in samples treated with IL-10, containing two or more functionally linked genes. The detection of these gene groups in a single operon under the influence of IL-8 and IL-10 suggests that in response to these cytokines, bacteria may regulate key biochemical pathways. This could have significant implications for their metabolism, stress response, and interaction with the host immune system. In turn, this may contribute to bacterial adaptation to the inflammatory environment or modulation of the inflammatory process.

Conclusions: The results demonstrate how the exposure to anti-inflammatory and pro-inflammatory cytokines alters the expression profiles in strains of *L. rhamnosus* K32. The reduction in the activity of genes involved in energy processes, such as carbohydrate exchange and transport, stimulates the transition of cells to a mode of minimal interaction with the external environment. This adaptive transition facilitates the conservation and redistribution of resources to maintain cell viability and activates protective mechanisms in conditions of inflammatory stress.

Funding: Financial support for this study was provided by the Russian Science Foundation under the grant number 22-75-10029, available at https://rscf.ru/project/22-75-10029/.

doi 10.18699/bgrs2024-1.4-10

База знаний RatDEGdb по дифференциально экспрессирующимся генам крысы как модельного объекта биомедицинских исследований

Пономаренко М.[1]*, Чадаева И.[1], Филонов С.[1], Золотарева К.[1], Кожемякина Р.[1], Подколодный Н.[1, 3], Хандаев Б.[1, 2], Рассказов Д.[1], Кондратюк Е.[1, 4, 5], Климова Н.[1], Федосеева Л.[1], Богомолов А.[1, 2], Кожевникова О.[1], Шихевич С.[1], Рязанова М.[1], Ощепков Д.[1, 2], Ершов Н.[1], Редина О.[1], Стефанова Н.[1], Колосова Н.[1], Маркель А.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Институт вычислительной математики и математической геофизики СО РАН, Новосибирск, Россия*

[4] *Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия*

[5] *Сибирский федеральный научный центр агробиотехнологий РАН, Краснообск, Новосибирская область, Россия*

* *pon@bionet.nsc.ru*

Ключевые слова: база знаний; ДЭГ; крысы *Rattus norvegicus*; животные модели болезней человека; RNA-seq; ПЦР; микрочипы

Мотивация и цель: Животные модели, используемые в биомедицинских исследованиях, в настоящее время охватывают практически весь известный спектр заболеваний человека. Они необходимы для понимания физиологических, генетических и эпигенетических механизмов, регулирующих эволюционно закрепленные фенотипические признаки организма, должны максимально повторять симптомы изучаемой патологии и соответствовать строгим критериям [1]. Чаще всего в этих исследованиях используют крыс и мышей, лабораторные линии которых насчитывают на сегодняшний день десятки тысяч [2].

Методы и алгоритмы: Нами создана база знаний RatDEGdb по дифференциально экспрессирующимся генам (ДЭГ) крысы в качестве одного из модельных объектов в биомедицинских исследованиях, которая представляет собой коллекцию ранее опубликованных экспериментальных данных по экспрессии генов у крыс разных линий, предназначенных для изучения артериальной гипертонии, возрастных нарушений, психопатологических состояний и других заболеваний человека.

Результаты: Текущий выпуск RatDEGdb содержит 25 101 ДЭГ, представляющих 14 320 уникальных генов крысы, которые изменяют уровень транскрипции в 21 ткани 10 генетических линий крысы в качестве моделей 11 заболеваний человека согласно 45 оригинальным научным статьям, как это охарактеризовано в табл. 1. Новшество RatDEGdb по сравнению с другими биомедицинскими базами данных – это курируемая аннотации отклонений ДЭГ крысы как модельного объекта с использованием независимых клинических данных об однонаправленных изменениях экспрессии гомологичных генов, выявленных у людей при различных патологиях. Собранные ДЭГ крыс были аннотированы однонаправленными изменениями экспрессии гомологичных им генов человека у

больных в сравнении со здоровыми. Текущий выпуск RatDEGdb содержит 94 873 такие аннотации для 321 гена человека при 836 заболеваниях согласно 959 оригинальным научным статьям, найденным в текущем выпуске базы данных PubMed [3].

Таблица 1. Примеры дифференциально экспрессируемых генов (ДЭГ) крысы как модельных животных в биомедицине, которые были выявлены с использованием технологий ПЦР, RNA-Seq или микрочипов, после чего их публикации были документированы в базе знаний RatDEGdb (https://www.sysbio.ru/RatDEGdb)

№	Линия	Ткань	Синдром	Модель	Норма	N	тип	[Лит]
1	Агрессивные	гипоталамус	Агрессия	Агрессивные	Ручные	1	ПЦР	[4]
2		лобная кора	Агрессия	Агрессивные	Ручные	1		[5]
3		средний мозг	Агрессия	Агрессивные	Ручные	1		[6]
4	SD	мозг	Агрессия	Агрессивные	Ручные	5		[7]
5	Ручные	гиппокамп	Агрессия	метил-богатая диета матери	норма	3		[8]
…	…	…	…	…	…	…		…
25	ГК	средний мозг	Каттония	ГК	WAG	1		[9]
26	НИСАГ	Почка	гипертония	НИСАГ	WAG	1		[10]
27	OXYS	Сетчатка	возрастная макулярная дегенерация	OXYS	Wistar	5		[11]
28	Агрессивные	лобная кора	Агрессия	Агрессивные	Ручные	24	RNA-seq	[12]
29		гипоталамус	Агрессия	Агрессивные	Ручные	46		[13]
30		гиппокамп	Агрессия	Агрессивные	Ручные	42		[14]
…	…	…	…	…	…	…		…
63	НИСАГ	надпочечник	гипертония	НИСАГ	WAG	1020		[15]
64	OXYS	префронтальная кора	болезнь Альцгеймера	возраст 5 месяцев	возраст 20 дне	52		[16]
65	Wistar	надпочечник	гипертония	дексаметазон, инъекция беременной	норма	93	микрочип	[17]
…	…	…	…	…	…	…		…
73	SHRSP	мозг	инсульт	SHRSP, 3 недели	SHR 3 недели	17		[18]
∑	10 линий	21 ткань	11 синдромов	73 модели заболеваний		25101	45 статей	

Примечание. N – количество ДЭГ; линии крыс: ГК – генетическая каттония; НИСАГ – наследственная стресс-индуцируемая артериальная гипертония; OXYS – уникальная селекционная модель преждевременного старения и связанных с ним заболеваний; SD – Sprague Dawley; SHR – спонтанная гипертония; SHRSP – спонтанная гипертония с склонностью к инсульту; Wistar – самая первая инбредная линия одомашненных лабораторных крыс, созданная в Институте Вистар (США) в 1906 г., WAG – дочерняя линия Wistar Albino Glaxo; ∑ – ИТОГО.

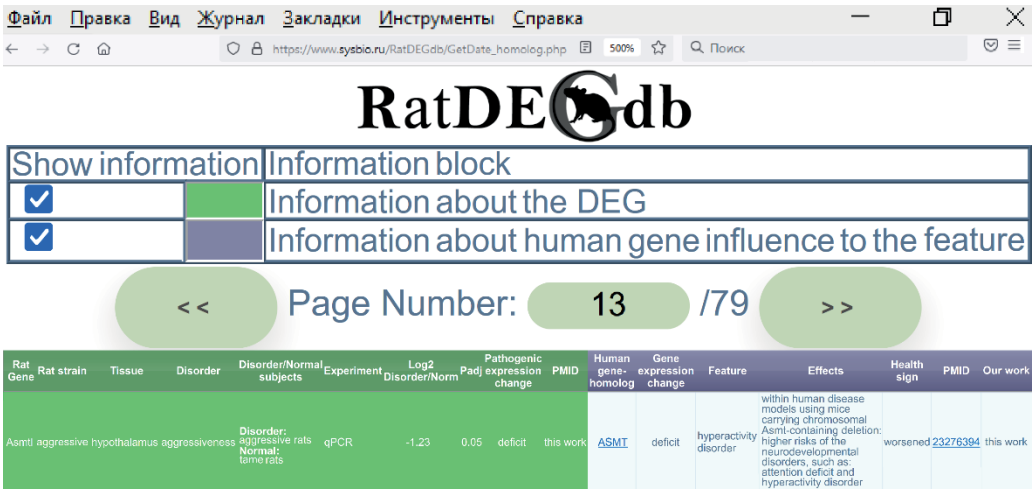


Рис. 1. Пример записи в базе знаний RatDEGdb, созданной в этой работе, документирующей экспериментальные данные о пониженном уровне экспрессии гена *Asmtl* в гипоталамусе крыс агрессивной линии в сравнении с ручной линией в качестве биомедицинской модели агрессивности как симптома заболеваний человека [4] вместе с их аннотацией (см. табл. 1, строка 1) с использованием независимых данных о низкой экспрессии гомологичного гена *ASMT* человека у больных с гиперактивностью в согласии с биомедицинской моделью этого заболевания с использованием мышей, несущих делецию гена Asmt, описанной в работе [19], индексированной в базе данных PubMed [3]: PMID: 23276394

Выводы: Представленная база знаний может быть интересна в первую очередь специалистам по генетике человека, молекулярным биологам, клиницистам и генетическим консультантам, а также специалистам в области биофармацевтики, биоинформатики и персонализированной геномики. RatDEGdb является свободно доступной (https://www.sysbio.ru/RatDEGdb), как показано на рис. 1.

Финансирование: Работа поддержана бюджетным проектом FWNR-2022-0020.

RatDEGdb: a knowledge base of differentially expressed genes in the rat as a model object in biomedical research

Ponomarenko M.[1]*, Chadaeva I.[1], Filonov S.[1, 2], Zolotareva K.[1], Kozhemyakina R.[1], Podkolodnyy N.[1, 3], Khandaev B.[1, 2], Rasskazov D.[1], Kondratyuk E.[1, 4, 5], Klimova N.[1], Fedoseeva L.[1], Bogomolov A.[1, 2], Kozhevnikova O.[1], Shikhevich S.[1], Ryazanova M.[1], Oshchepkov D.[1, 2], Ershov N.[1], Redina O.[1], Stefanova N.[1], Kolosova N.[1], Markel A.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

[4] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[5] *Siberian Federal Scientific Centre of Agro-BioTechnologies, RAS, Krasnoobsk, Novosibirsk Region, Russia*

* *pon@bionet.nsc.ru*

Key words: knowledge base; DEG; *Rattus norvegicus*; animal models of human diseases; RNA-seq; PCR; microarrays

Motivation and Aim: The animal models used in biomedical research cover virtually every human disease. The animal models required for understanding the physiological, genetic and epigenetic mechanisms regulating evolutionarily fixed phenotypic traits of an organism are supposed to perfectly mimic the symptoms of the pathology being studied and to conform to strict criteria [1]. The most popular animal models are rats and mice, with dozens of thousands of laboratory strains in use [2].

Table 1. Examples of differentially expressed genes (DEGs) of the rat as a biomedical model animal that were identified using qPCR, RNA-Seq or microarray technologies, and then their publications were documented in the knowledge base RatDEGdb presented in this work, which is public available (https://www.sysbio.ru/RatDEGdb)

#	Strain	Tissue	Syndrome	Model	Norm	N	Measure	[Ref]
1	Aggressive	hypothalamus	aggression	aggressive	tame	1	qPCR	[4]
2		frontal cortex	aggression	aggressive	tame	1		[5]
3		midbrain	aggression	aggressive	tame	1		[6]
4	SD	brain	aggression	aggressive	tame	5		[7]
5	Tame	hippocampus	aggression	maternal methyl-rich diet	norm	3		[8]
…	…	…	…	…	…	…		…
25	GK	midbrain	catatonia	GK	WAG	1		[9]
26	ISIAH	kidney	hypertension	ISIAH	WAG	1		[10]
27	OXYS	retina	age-related macular degeneration	OXYS	Wistar	5		[11]
28	Aggressive	frontal cortex	aggression	Aggressive	Tame	24	RNA-seq	[12]
29		hypothalamus	aggression	Aggressive	Tame	46		[13]
30		hippocampus	aggression	Aggressive	Tame	42		[14]
…	…	…	…	…	…	…		…
63	ISIAH	adrenal gland	hypertension	ISIAH	WAG	1020		[15]
64	OXYS	prefrontal cortex	Alzheimer's disease	5 months old	20 days old	52		[16]
65	Wistar	adrenal gland	hypertension	dexamethasone injection for pregnant rats	norm	93	micro array	[17]
…	…	…	…	…	…	…		…
73	SHRSP	brain	stroke	SHRSP	SHR	17		[18]
∑	10 strains	21 tissues	11 diseases	73 biomedical models		25101	45 articles	

Note. N – amount of DEGs; rat strains: GC – genetic catatonia; ISIAH – hereditary stress-induced arterial hypertension; OXYS – unique selection-based model of premature ageing and associated diseases; SD – Sprague Dawley; SHR – spontaneous hypertension; SHRSP – spontaneous hypertension with stroke tendency; Wistar – the very first inbred line of domesticated laboratory rats, which was created in 1906 at the Wistar Institute (USA), WAG is a daughter line of Wistar Albino Glaxo; ∑ – TOTAL.

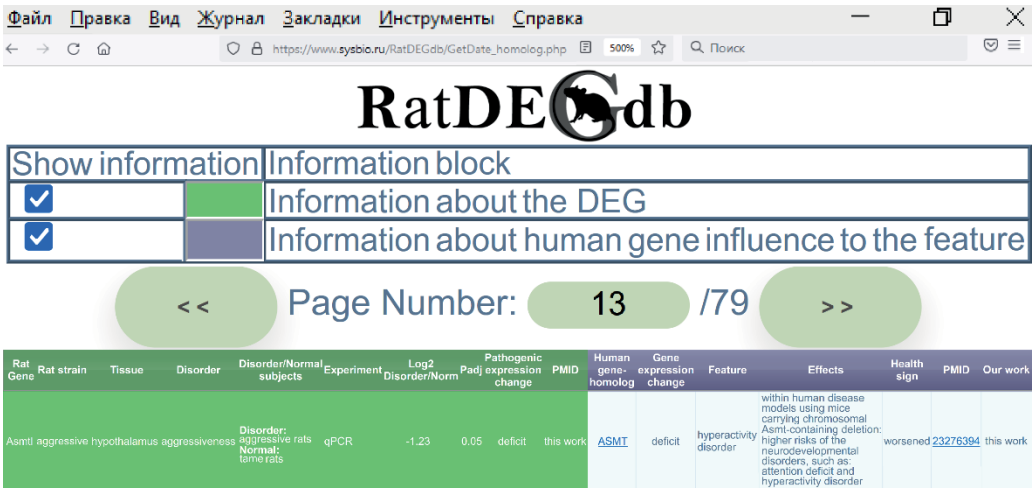


Fig. 1. A sample entry in RatDEGdb documents experimental data on *Asmtl* expression in the hypothalamus of aggressive rats compared to the tame rats as a biomedical model of aggressive behavior in human diseases [4] along with their annotation (Table 1: row #1) using independent data on deficit of its human homolog *ASMT* in patients with hyperactivity in line with an *Asmt*-null mouse model of this disease, as described by the article [19] indexed by the database PubMed [3]: PMID: 23276394)

Methods and Algorithms: We have created RatDEGdb, a knowledge base of the differentially expressed genes (DEGs) of the rat as a model object in biomedical research is a collection of published data on gene expression in rat strains simulating arterial hypertension, age-related disorders, psychopathological conditions and other diseases.
Results: The current release contains information on 25,101 DEGs representing 14,320 unique rat genes that change transcription levels in 21 tissues of 10 genetic rat strains used as models of 11 human diseases based on 45 original scientific papers, as described in the Table. RatDEGdb is novel in that, unlike any other biomedical database, it offers the manually curated annotations of DEGs in model rats with the use of independent clinical data on equal changes in the expression of homologous genes revealed in people with pathologies. The rat DEGs put in RatDEGdb were annotated with equal changes in the expression of their human homologs in affected people. In its current release, RatDEGdb contains 94,873 such annotations for 321 human genes in 836 diseases based on 959 original scientific papers found in the current PubMed [3].
Conclusion: RatDEGdb may be interesting to human geneticists, molecular biologists, clinicians, genetic advisors as well as experts in bioinformatics, personalized genomics and pharmaceutics. RatDEGdb is freely available (https://www.sysbio.ru/RatDEGdb), as shown on the Fig. 1.
Funding: The study is supported by the Government Budget Project FWNR-2022-0020.

Список литературы/References

1. Gryksa K. et al. Selective breeding of rats for high (HAB) and low (LAB) anxiety-related behaviour: a unique model for comorbid depression and social dysfunctions. *Neurosci Biobehav Rev.* 2023;152:105292
2. Гайдай Е.А., Гайдай Д.С. Генетическое разнообразие экспериментальных мышей и крыс: история возникновения, способы получения и контроля. *Лабораторные животные для научных исследований*. 2019;4:78-85

[Gajdaj E.A., Gajdaj D.S. Genetic diversity of experimental mice and rats: history of origin, methods of production and control. *Laboratory Animals for Scientific Research*. 2019;4:78-85 (in Russian)]
3. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)*. 2011;2011:baq036
4. Чадаева И. и др. База знаний RatDEGdb по дифференциально экспрессирующимся генам крысы как модельного объекта биомедицинских исследований. *Вавиловский журнал генетики и селекции*. 2023;27(7):794-806
[Chadaeva I. et al. RatDEGdb: a knowledge base of differentially expressed genes in the rat as a model object in biomedical research. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):794-806 (in Russian)]
5. Naumenko V. et al. Expression of serotonin transporter gene and startle response in rats with genetically determined fear-induced aggression. *Bull Exp Biol Med*. 2009;147(1):81-83
6. Popova N. et al. Involvement of brain serotonin 5-HT1A receptors in genetic predisposition to aggressive behavior. *Neurosci Behav Physiol*. 2007;37(6):631-635
7. Suzuki H. et al. Increased 5-HT1B receptor density in the basolateral amygdala of passive observer rats exposed to aggression. *Brain Res Bull*. 2010;83(1-2):38-43
8. Herbeck Yu. et al. Effects of maternal methyl-supplemented diet on hippocampal glucocorticoid receptor mRNA expression in rats selected for behavior. *Cytol Genet (Kyiv)*. 2010;44(2):108-113
9. Рязанова М. и др. Экспрессия генов системы катехоламинов в среднем мозге и реакция престимульного торможения у крыс с генетической кататонией. *Вавиловский журнал генетики и селекции*. 2017;21(7):798-803
[Ryazanova M. et al. Expression of catecholaminergic genes in the midbrain and prepulse inhibition in rats with a genetic catatonia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2017;21(7):798-803 (in Russian)]
10. Fedoseeva L. et al. Renin system of the kidney in ISIAH rats with inherited stress-induced arterial hypertension. *Bull Exp Biol Med*. 2009;147(2):177-180
11. Perepechaeva M. et al. The mitochondria-targeted antioxidant SkQ1 downregulates aryl hydrocarbon receptor-dependent genes in the retina of OXYS rats with AMD-like retinopathy. *J Ophthalmol* 2014;2014:530943
12. Albert F. et al. A comparison of brain gene expression levels in domesticated and wild animals. *PLoS Genet*. 2012;8(9):e1002962
13. Chadaeva I. et al. Domestication explains two-thirds of differential-gene-expression variance between domestic and wild animals; the remaining one-third reflects intraspecific and interspecific variation. *Animals*. 2021;11(9):2667
14. Oshchepkov D. et al. Stress reactivity, susceptibility to hypertension, and differential expression of genes in hypertensive compared to normotensive patients. *Int J Mol Sci*. 2022a;23(5):2835
15. Fedoseeva L. et al. Molecular determinants of the adrenal gland functioning related to stress-sensitive hypertension in ISIAH rats. *BMC Genomics*. 2016;17(Suppl. 14):989
16. Stefanova N.A. et al. The rat prefrontal-cortex transcriptome: effects of aging and sporadic Alzheimer's disease-like pathology. *J Gerontol A Biol Sci Med Sci*. 2019;74(1):33-43
17. Tharmalingam S. et al. Whole transcriptome analysis of adrenal glands from prenatal glucocorticoid programmed hypertensive rodents. *Sci Rep*. 2020;10(1):18755
18. Yoshida M. et al. Analysis of genes causing hypertension and stroke in spontaneously hypertensive rats: gene expression profiles in the brain. *Int J Mol Med*. 2014;33(4):887-896
19. Trent S. et al. Biological mechanisms associated with increased perseveration and hyperactivity in a genetic mouse model of neurodevelopmental disorder. *Psychoneuroendocrinology*. 2013;38(8):1370-1380

doi 10.18699/bgrs2024-1.4-11

Индуцируемое гипокинезией изменение транскриптома различается в разных скелетных мышцах: общие и уникальные механизмы

Попов Д.В.*, Курочкина Н.С., Жедяев Р.Ю., Борзых А.А., Леднев Е.М., Вепхвадзе Т.Ф., Шпаков А.В.

Институт медико-биологических проблем РАН, Москва, Россия

** danil-popov@yandex.ru*

Ключевые слова: гипокинезия; транскриптом; скелетная мышца

Мотивация и цель: Скелетная мускулатура составляет более трети от массы тела человека; при нормальном уровне двигательной активности скелетные мышцы играют ключевую роль в регуляции углеводно-жирового обмена и инсулиновой чувствительности организма. Хроническое снижение двигательной активности (гипокинезия) вызывает снижение способности мышц окислять жиры и углеводы, понижает инсулиновую чувствительность, содержание митохондрий в них и их массу, что оказывает негативное влияние на метаболизм и работоспособность человека. Интересно, что выраженность негативных эффектов гипокинезии различна для разных мышц (например, камбаловидной мышцы голени (*m. soleus*) и четырехглавой мышцы бедра (*m. vastus lateralis*)), однако молекулярные механизмы, объясняющие эти различая, изучены явно недостаточно. Цель работы – сопоставить изменения транскриптома в мышцах бедра и голени после трехнедельной гипокинезии и выявить гены, ассоциированные с общим и уникальным ответом на гипокинезию.

Методы и алгоритмы: В эксперименте участвовали молодые (25–38 лет, $n = 12$) здоровые мужчины. Исследовали влияние 3-недельной постельной гипокинезии на силу и работоспособность (выносливость) мышц-разгибателей коленного сустава и сгибателей голеностопного сустава. До и после гипокинезии брали биопсические пробы ткани из *m. soleus* и *m. vastus lateralis* для исследования экспрессии белок-кодирующих генов (РНК-секвенирование, одноконцевые прочтения, 66 млн прочтений/образец, NextSeq 550, Illumina).

Результаты: Гипокинезия привела к выраженному и сопоставимому снижению силы мышц обеих мышечных групп на 12 %, тогда как снижение выносливости было обнаружено только для мышц голени (10 %). После гипокинезии в *m. soleus* изменили экспрессию в три раза больше генов, чем в *m. vastus lateralis* (более 600 и 1500 мРНК соответственно); эти изменения были связаны преимущественно со снижением экспрессии. Интересно, что две трети генов, изменивших экспрессию в *m. vastus lateralis*, пересекались с изменениями в *m. soleus*, т. е. были «общими» для обеих мышц. Среди них гены, увеличившие экспрессию, обогатили функциональную категорию «клеточная мембрана» и включали ряд транспортеров, тогда как гены, снизившие экспрессию, были ассоциированы с митохондриальными дыхательными ферментами, регуляторами жирового обмена и саркомерными белками Z-диска. Многочисленные гены, «уникальные» для *m. soleus*, обогатили сходные функциональные категории. Отличительной

особенностью «уникального» ответа этой мышцы было то, что среди генов, снизивших экспрессию, было более 200 генов, кодирующих различные белки митохондрий, а гены, увеличившие экспрессию, были ассоциированы с различными провоспалительными сигнальными каскадами (включая сигнальный путь TNF) и секретируемыми факторами.
Выводы: Наши результаты показывают, что различия в изменении функциональных возможностей мышц бедра и голени после 3-недельной гипокинезии коррелируют с изменениями их транскриптомного профиля. Выявление общих и уникальных механизмов, лежащих в основе ответа разных мышц, открывает перспективы для поиска фармакологических подходов для таргетной профилактики снижения функциональных возможностей разных мышц при гипокинезии.

Disuse-induced transcriptomic changes vary in different skeletal muscles: common and unique mechanisms

Popov D.V.*, Kurochkina N.S., Zhedyaev R.Yu., Borzykh A.A., Lednev E.M., Vepkhvadze T.F., Shpakov A.V.
Institute of Biomedical Problems, RAS, Moscow, Russia
** danil-popov@yandex.ru*

Key words: disuse; transcriptome; skeletal muscle

Motivation and Aim: Skeletal muscle makes up more than a third of human body mass; with a normal level of physical activity, skeletal muscles play a key role in the regulation of carbohydrate and fat metabolism and insulin sensitivity of the organism. Chronic decrease in motor activity (disuse) causes a decrease in the ability of muscles to oxidize fats and carbohydrates, reduces insulin sensitivity, the content of mitochondria and muscle mass, which has a negative impact on human metabolism and performance. Interestingly, the severity of the negative effects of disuse varies for different muscles (for example, the soleus muscle and the vastus lateralis muscle), but the molecular mechanisms that explain these differences are clearly not well understood. The purpose of the work is to compare transcriptomic changes in the muscles of the thigh and lower leg after three weeks of disuse and to identify genes associated with a common and unique response to disuse.
Methods and Algorithms: Young (25–38 years old, $n = 12$) healthy men participated in the experiment. The effect of 3 weeks of bed rest on the strength and performance (endurance) of the knee extensor and ankle flexor muscles was studied. Before and after disuse, biopsy tissue samples were taken from *m. soleus* and *m. vastus lateralis* to study the expression of protein-coding genes (RNA-sequencing, single-end reads, 66 million reads/sample, NextSeq 550, Illumina).
Results: Disuse led to a pronounced and comparable decrease in muscle strength of both muscle groups by 12 %, while a decrease in endurance was found only for the lower leg muscles (10 %). After disuse in *m. soleus* changed the expression of 3 times more genes than in *m. vastus lateralis* (more than 600 and 1500 mRNAs, respectively); these changes were associated primarily with decreased expression. Interestingly, two-thirds of the genes that changed expression in *m. vastus lateralis* intersected with changes in

m. soleus, i. e. were “common” for both muscles. Among them, genes that increased expression enriched the functional category “cell membrane” and included a number of transporters, while genes that decreased expression were associated with mitochondrial respiratory enzymes, lipid metabolism regulators, and sarcomeric Z-disk proteins. Numerous genes “unique” to *m. soleus*, enriched similar functional categories. A distinctive feature of the “unique” response of this muscle was that among the down-regulated genes were more than 200 genes encoding various mitochondrial proteins, while the up-regulated genes were associated with various pro-inflammatory signaling cascades (including the TNF signaling pathway) and secreted factors.

Conclusion: Our results show that differences in changes in the functional capacity of the thigh and lower leg muscles after 3 weeks of disuse correlate with changes in their transcriptomic profile. Identification of common and unique mechanisms underlying the response of different muscles opens up prospects for the search for pharmacological approaches for targeted prevention of decreased functionality of different muscles during disuse.

doi 10.18699/bgrs2024-1.4-12

Молекулярные механизмы ответа клеточной стенки ячменя на различные типы ионизирующего излучения

Празян А.[1]*, Подлуцкий М.[1], Бондаренко Е.[1], Волкова П.[2], Казакова Е.[1], Битаришвили С.[1], Шестерикова Е.[1], Макаренко Е.[1], Лыченкова М.[1], Гераськин С.[1], Король М.[1]

[1] *Всероссийский научно-исследовательский институт радиологии и агроэкологии Национального исследовательского центра «Курчатовский институт», Обнинск, Россия*

[2] *Независимый исследователь, Гел, Бельгия*

* *prazyana@yahoo.com*

Ключевые слова: ячмень; радиация; DEG; электрон; протон; гамма

Мотивация и цель: Повышение урожайности сельскохозяйственных культур имеет решающее значение для современного мира, где обеспечение продовольственной безопасности является первостепенной задачей. Увеличение частоты экстремальных климатических условий создает дополнительную нагрузку на сельскохозяйственные растения, что требует разработки стратегий адаптации [1]. Ионизирующее излучение (ИИ) дает представление о механизмах адаптации растений к абиотическим стрессорам, активируя антиоксидантную защиту и механизмы репарации ДНК [2, 3]. Ячмень (*Hordeum vulgare* L.) – важный сельскохозяйственный вид, является потенциальным объектом для повышения урожайности и улучшения устойчивости. Радиобиологические исследования реакций ячменя выявляют молекулярные пути, участвующие в адаптации к абиотическим стрессам [4–6]. Было проведено расширенное мультиомиксное исследование с использованием ионизирующего излучения с низким и высоким ЛПЭ для обработки семян озимого ячменя [7]. Полный сравнительный анализ общего эффекта трех типов ИИ на транскриптомный профиль ячменя, а также дополнительные материалы представлены в [8].

В нашем исследовании мы использовали гамма-, электронное и протонное излучение в сочетании с секвенированием всего транскриптома, чтобы идентифицировать общие ответные транскрипты и оценить их роль в реакциях на стресс.

Методы и алгоритмы: Семена озимого ячменя сорта Фокс 1 были посажены в горшки с почвой и пророщены в контролируемых условиях. Сеянцы были облучены на 7-й день после посева, а листья были собраны через 24 часа после облучения. Растения из одного горшка были объединены в одну пробу.

Проростки подвергались воздействию γ-излучения, электронов и протонов в дозе 15 Гр с использованием источника γ-излучения «Агат» (изотоп ^{60}Co), линейного ускорителя электронов NOVAC11 и комплекса протонной терапии «Прометей». Дозиметрия проводилась в соответствии с установленными протоколами для обеспечения точного облучения.

Подготовка библиотек и секвенирование были проведены компанией «Евроген» (Москва, Россия) на Illumina NovaSeq 6000 (2 × 150 пар оснований). Обработка данных включала контроль качества, фильтрацию низкокачественных прочтений и выравнивание прочтений с использованием эталонного транскриптома ячменя (BaRTv2.18). Дифференциальная экспрессия генов анализировалась с

использованием DESeq2 для выявления генов, экспрессия которых значимо различалась между контрольными и облученными образцами. Полученные данные аннотировались и анализировались с использованием баз данных UniProt и NCBI для получения дополнительной информации о белках и их функциях. Для лучшего понимания биологических процессов, затронутых ионизирующим излучением, был проведен функциональный анализ обогащения с использованием базы данных Gene Ontology. Этот анализ позволил выявить группы генов, ассоциированные с определенными биологическими процессами, молекулярными функциями и клеточными компонентами, которые значительно изменились после облучения.

Результаты: Дифференциальный анализ экспрессии генов в растениях после воздействия гамма-, протонного и электронного излучения выявил 553 уникальных гена с повышенной экспрессией и 124 гена с пониженной экспрессией. При сравнении общих транскрипционных ответов на три типа облучения было выявлено 47 общих ДЭГ. Ионизирующее излучение воздействует на клетки растений, превращая нейтральные атомы или молекулы в активное состояние. Этот процесс вызывает различные биологические эффекты, включая модуляцию систем сигнальной передачи растений и индукцию окислительного стресса за счет увеличения продукции активных форм кислорода (АФК). Наше исследование подчеркивает значительное влияние трех типов ИИ на ячмень на молекулярном уровне, особенно выражающееся в повышении экспрессии генов, связанных с биосинтезом и модификацией клеточной стенки, таких как синтазы целлюлозы (*D9IXC7* и *F2DMG1*) и бета-галактозидазы (*A0A8I6X4I5*). Синтазы целлюлозы критически важны для производства целлюлозы, основного компонента клеточной стенки, в то время как бета-галактозидазы участвуют в модификации и ремоделировании полисахаридов клеточной стенки. Эти изменения предполагают адаптивный ответ для укрепления целостности и функции клеточной стенки под воздействием стресса. Два гена лакказ (*F2DXF2* и *F2D9J2*) имели повышенную экспрессию при всех типах ИИ. Лакказы являются оксидоредуктазными ферментами, которые катализируют окислительные реакции с участием ароматических и неароматических соединений, играя важную роль в функционировании клеточной стенки и биодеградации лигнина. Ранее лакказы ассоциировались с улучшением устойчивости к засухе за счет повышения водотранспортной способности и усиления антиоксидантных ответов у других видов растений. Ген *M0XEM7* также имел повышенную экспрессию. Этот ген важен для прорастания пыльцы и роста пыльцевых трубок, что включает биосинтез и модификацию клеточной стенки, указывая на его роль в поддержании целостности клеточной стенки при стрессе, вызванном ИИ. Кроме того, имел повышенную экспрессию ген, кодирующий белок, похожий на гермин 5-1 (GLP, кодируемый *A0A287KQB5*), известный своей ролью в накоплении лигнина и уменьшении содержания H_2O_2 через активность ферментов, связанных с АФК, что указывает на двойную роль в укреплении клеточных стенок и управлении окислительным стрессом. Мы также отметили повышение экспрессии генов, связанных с метаболизмом фитогормонов, особенно тех, которые связаны с биосинтезом и сигнальной передачей ауксинов (*A0A8I6YN81* и *A0A8I6XJC3*). Ауксины важны для регуляции ответов растений на стресс, включая модификацию клеточной стенки. Известно, что экспрессия синтаз целлюлозы и бета-галактозидаз регулируется фитогормонами, что позволяет предположить, что изменения

в балансе фитогормонов, вызванные ИИ, могут модулировать биосинтез и ремоделирование клеточной стенки в рамках адаптивного ответа растений.
Выводы: Клеточная стенка, вероятно, является одной из основных целей облучения γ-, электронами и протонами, что характеризуется значительным увеличением экспрессии генов, участвующих в ее биосинтезе, модификации и укреплении. Этот защитный ответ на уровне клеточной стенки, скорее всего, служит для поддержания структурной целостности и управления оксидативным стрессом, вызванным ИИ. Понимание молекулярных механизмов может способствовать разработке стратегий увеличения устойчивости растений к абиотическим стрессам, что потенциально может улучшить урожайность и качество урожая в неблагоприятных климатических условиях, так как обнаруженные общие реагирующие молекулы также участвуют в ответах на другие абиотические стрессовые факторы (засуха, соленость, воздействие тяжелых металлов).
Благодарности: Ячменные зерна были любезно предоставлены Российским научным учреждением «Научно-исследовательский центр сельского хозяйства «Донской». Мы также выражаем нашу благодарность Медицинскому радиологическому исследовательскому центру им. А.Т. Цыба за облучение наших растений различными типами ионизирующего излучения.
Финансирование: Эта работа была финансово поддержана Министерством науки и высшего образования Российской Федерации (Договор № 075-15-2021-1068 от 28 сентября 2021 г.).

Molecular mechanisms of barley cell wall response to different types of ionising radiation

Prazyan A.[1]*, Podlutskii M.[1], Bondarenko E.[1], Volkova P.[2], Kazakova E.[1], Bitarishvili S.[1], Shesterikova E.[1], Makarenko E.[1], Lychenkova M.[1], Geras'kin S.[1], Korol M.[1]

[1] *Russian Institute of Radiology and Agroecology of National Research Centre "Kurchatov Institute", Obninsk, Russia*
[2] *Independent Researcher, Geel, Belgium*
* *prazyana@yahoo.com*

Key words: barley; radiation; DEG; electron; proton; gamma

Motivation and Aim: Enhancing crop yield is crucial for the modern world, where ensuring food security is paramount. The increasing frequency of extreme climatic conditions poses additional stress on agricultural plants, necessitating the development of adaptation strategies [1]. Ionizing radiation (IR) offers insights into plant adaptation mechanisms to abiotic stressors, activating antioxidant defense and DNA repair mechanisms [2, 3]. Barley (*Hordeum vulgare* L.), a significant agricultural species, shows potential for yield enhancement and resilience improvement. Radiobiological studies of barley responses reveal molecular pathways involved in abiotic stress adaptation [4–6]. An extended multiomics experiment involving ionising radiation with low- and high-LET exposure of winter barley seedlings was performed [7]. The full comparative analysis of the common effect of the three types of IR on transcriptomic profile of barley, along with additional materials, are presented in [8].

In our study we employed gamma, electron, and proton radiation, coupled with whole-transcriptome sequencing, to identify commonly responsive transcripts and assess their role in stress responses.

Methods and Algorithms: Winter barley cultivar Fox 1 seeds were planted in soil-filled pots and germinated under controlled conditions. Seedlings were irradiated on the 7th day after sowing, and leaves were harvested 24 hours post-irradiation. Plants from the same pot were pooled into one sample.

Seedlings were exposed to γ-radiation, electrons, and protons at a dose of 15 Gy using γ-radiation source "Agat" (^{60}Co isotope), NOVAC11 linear electron accelerator, and Prometheus Proton Therapy Complex. Dosimetry was performed according to established protocols to ensure accurate exposure.

Library preparation, and sequencing were provided by the "Evrogen" (Moscow, Russia) at Illumina NovaSeq 6000 (2 × 150 bp). Data processing involved quality control, filtering of low-quality reads, and alignment of sequenced reads using the barley reference transcriptome (BaRTv2.18). Differential gene expression was analyzed using DESeq2 to identify genes, the expression of which significantly differed between control and irradiated samples. The obtained data was annotated and analyzed using UniProt and NCBI databases to obtain additional information about the proteins and their functions. For a better understanding of the biological processes affected by ionizing radiation, functional enrichment analysis was performed using the Gene Ontology database. This analysis allowed the identification of gene groups associated with specific biological processes, molecular functions, and cellular components that were significantly altered after irradiation.

Results: The differential gene expression analysis in plants following exposure to gamma, proton, and electron radiation, revealed 553 unique genes with increased expression and 124 genes with decreased expression. When comparing transcriptional responses to the three types of irradiations, 47 common DEGs were identified. Ionising radiation affects the plant cells, converting neutral atoms or molecules into reactive ions. This process triggers various biological effects, including the modulation of plant signalling systems and induction of oxidative stress via increased production of reactive oxygen species (ROS). Our study highlights the significant impact of the three types of IR on barley at the molecular level, particularly upregulating genes associated with the cell wall biosynthesis and modification, such as cellulose synthases (*D9IXC7* and *F2DMG1*) and beta-galactosidase (*A0A8I6X4I5*). Cellulose synthases are critical for the production of cellulose, a major component of the cell wall, while beta-galactosidases are involved in the modification and remodeling of cell wall polysaccharides. These changes suggest an adaptive response to reinforce cell wall integrity and function under stress. Two laccase genes (*F2DXF2* and *F2D9J2*) were upregulated across all types of IR exposure. Laccases are multicopper oxidase enzymes that catalyse oxidative reactions involving aromatic and non-aromatic compounds, playing a vital role in cell wall functioning and lignin biodegradation. Laccases have been previously associated with improving drought tolerance by enhancing water transport capacity and strengthening antioxidant responses in other plant species. The gene *M0XEM7*, encoding a Cd^{2+}-induced protein (nucleotide-diphospho-sugar transferase), was also upregulated. This gene is essential for pollen germination and tube growth, which involves cell wall biosynthesis and modification, indicating its role in maintaining cell wall integrity under IR-induced stress. Additionally, germin-like protein 5-1 (GLP, encoded by *A0A287KQB5*), known for its involvement in lignin accumulation and reduction of H_2O_2

content through ROS-related enzyme activity, was upregulated. This suggests a dual role in reinforcing cell walls and managing oxidative stress. We have also noted the upregulation of genes involved in phytohormone metabolism, particularly those related to auxin biosynthesis and signalling (*A0A8I6YN81* and *A0A8I6XJC3*). Auxins are crucial for regulating plant stress responses, including cell wall modifications. The expression of cellulose synthases and beta-galactosidases is known to be influenced by phytohormones, suggesting that IR-induced changes in phytohormone balance may modulate cell wall biosynthesis and remodeling as part of the plant's adaptive response.

Conclusion: The cell wall appears to be one of the main targets of γ-, electrons, and protons irradiation, characterized by significant upregulation of genes involved in its biosynthesis, modification, and reinforcement. This protective response at the cell wall level most probably serves to maintain structural integrity and manage the oxidative stress induced by IR. Understanding the molecular mechanisms can contribute to developing strategies to enhance plant resilience to abiotic stresses, potentially improving crop yield and quality in adverse environmental conditions, as the revealed commonly responsive molecules are also involved in responses to other abiotic stress factors (drought, salinity, exposure to heavy metals).

Acknowledgements: The barley grains were kindly provided by the Russian State Scientific Establishment "Agricultural Research Center "Donskoy"". And we would like to express our gratitude to the A. Tsyb Medical Radiological Research Centre in Obninsk, Russia, for irradiating our plants with various types of ionizing radiation.

Funding: This work was financially supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-1068 of September 28, 2021).

Список литературы/References

1. Cordell D., Drangert J.O., White S. The story of phosphorus: global food security and food for thought. *Global Environ Change*. 2009;19(2):292-305
2. Duarte G.T., Volkova P.Y., Fiengo Perez F., Horemans N. Chronic ionizing radiation of plants: an evolutionary factor from direct damage to non-target effects. *Plants*. 2023;12(5):1178. doi 10.3390/plants12051178
3. Podlutskii M., Babina D., Podobed M., Bondarenko E., Bitarishvili S., Blinova Y., Shesterikova E., Prazyan A. et al. *Arabidopsis thaliana* accessions from the Chernobyl exclusion zone show decreased sensitivity to additional acute irradiation. *Plants*. 2022;11(22):3142. doi 10.3390/plants11223142
4. Geras'kin S., Churyukin R., Volkova P. Radiation exposure of barley seeds can modify the early stages of plants' development. *J Environ Radioact.* 2017;177:71-83. doi 10.1016/j.jenvrad.2017.06.008
5. Kim J.H., Ryu T.H., Lee S.S., Lee S., Chung B.Y. Ionizing radiation manifesting DNA damage response in plants: An overview of DNA damage signaling and repair mechanisms in plants. *Plant Sci.* 2019;278:44-53. doi 10.1016/j.plantsci.2018.10.013
6. Ali H., Ghori Z., Sheikh S., Gul A. Effects of Gamma Radiation on Crop Production. In: Hakeem K. (Ed.). Crop Production and Global Environmental Issues. Springer, 2015:27-78. doi 10.1007/978-3-319-23162-4_2
7. Volkova P., Prazyan A., Podlutskii M. et al. Multi-omics responses of barley seedlings to low and high linear energy transfer irradiation. *Environ Exp Bot*. 2024;218:105600. doi 10.1016/j.envexpbot.2023.105600
8. Prazyan A., Podlutskii M., Volkova P. et al. Comparative analysis of the effect of gamma-, electron, and proton irradiation on transcriptomic profile of hordeum vulgare L. seedlings: in search for molecular contributors to abiotic stress resilience. *Plants*. 2024;13(3):342. doi 10.3390/plants13030342

doi 10.18699/bgrs2024-1.4-13

Предсказание и анализ длинных некодирующих РНК на основе пантранскриптома

Пронозин А.[1, 2]*, Афонников Д.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

** pronozinartem95@gmail.com*

Ключевые слова: днРНК; пантранскриптом; кукуруза

Мотивация и цели: Длинные некодирующие РНК (днРНК) представляют собой класс линейных или кольцевых молекул РНК длиной от 200 нуклеотидов, не кодирующих белок. К настоящему времени идентифицировано более полумиллиона последовательностей днРНК для различных организмов.

Однако, несмотря на растущее число исследований, структурно-функциональные характеристики днРНК по-прежнему остаются малоизученными. Это связано со множеством факторов, которые нужно учитывать при идентификации днРНК. Размер днРНК схож с размером некоторых генов, кодирующих белок, что приводит к ошибкам перепредсказания. Также исследования показывают, что последовательности днРНК претерпевают быструю эволюцию, закономерности которой пока не исследованы. Таким образом, возникает необходимость анализа генного состава сразу у нескольких представителей вида. Для этих целей были предложены концепции пангенома и пантранскриптома.

Однако на сегодняшний день работы, посвященные исследованию пантранскриптомов, в основном направлены на выявление и исследование новых генов, кодирующих белок. Тогда как работ, посвященных исследованию днРНК в масштабах пантранскриптомов, не так много, в особенности для растений.

Данная работа направлена на расширение знаний о пантранскриптоме кукурузы за счет дополнительного анализа днРНК.

Методы и алгоритмы: В работе использовалось 503 библиотеки инбредных линий кукурузы, полученных в работе [1]. Для сборки и анализа пантранскриптома был разработан биоинформатический конвейер, который включает три этапа обработки: контроль качества и сборка транскриптома, разделение транскриптомов на кодирующую и некодирующую части и анализ пантранскриптомов.

Результаты: В результате получено два пантранскриптома кукурузы. Пантранскриптом, содержащий кодирующую часть, насчитывает 33023 гена, кодирующих белок. Пантранскриптом, содержащий некодирующую часть, насчитывает 2215493 днРНК. Для обоих пантранскриптомов была проведена аннотация, определены группы ортологов, охарактеризованы консервативные и вариабельные части. Также отдельно для пантранскриптома с некодирующей частью был проведен поиск новых днРНК путем выравнивания полученных транскриптов с базами данных известных днРНК и анализа экспрессии выявленных новых днРНК. Для пантранскриптома с кодирующей частью был проведен анализ GO-терминов и экспрессии выявленных новых генов, кодирующих белок.

Выводы: Получены два пантранскриптома кукурузы, содержащие кодирующую (генов кодирующих белок) и некодирующую (днРНК) части. Для обоих пантранскриптомов были выявлены группы ортологов, а также консервативные и вариабельные части. Кроме того, для пантранскриптома, включающего некодирующие регионы, был проведен поиск новых днРНК.
Финансирование: Работа выполнена за счет финансирования Курчатовского геномного центра ФИЦ ИЦиГ СО РАН, соглашение с Министерством образования и науки РФ № 075-15-2019-1662. Вычисления проводились с использованием ресурсов ЦКП «Биоинформатика».

Pan-transcriptome-based long non-coding RNAs prediction and analysis

Pronozin A.[1, 2]*, Afonnikov D.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *pronozinartem95@gmail.com*

Key words: lncRNA; pan-transcriptome; maize

Motivation and Aim: Long non-coding RNAs (lncRNA) are a class of linear or circular RNA molecules of 200 nucleotides or more in length that do not encode proteins. To date, more than half a million lncRNA sequences have been identified for various organisms.
However, despite the growing number of studies, the structural and functional characteristics of lncRNAs remain poorly understood. This is due to the many factors that need to be considered when identifying lncRNAs. The size of lncRNAs is similar to that of some protein-coding genes, leading to overprediction errors. Also, studies show that lncRNA sequences undergo rapid evolution, the patterns of which have not yet been investigated. Thus, there is a need to analyze the gene composition in several members of a species at once. For this purpose, the concepts of pan-genome and pan-transcriptome have been proposed.
However, to date, the works devoted to the study of pan-transcriptoms are mainly focused on the identification and study of new protein-coding genes. Whereas, there are not many works investigating lncRNAs on the scale of pan-transcriptoms, especially for plants.
This work aims to expand the knowledge of the maize pan-transcriptome by further analyzing lncRNAs.
Methods and Algorithms: 503 libraries of maize inbred lines obtained in [1] were used in this work. A bioinformatics pipeline was developed for pan-transcriptome assembly and analysis, which includes three processing steps: quality control and transcriptome assembly, separation of transcriptomes into coding and non-coding parts, and pan-transcriptome analysis.
Results: Two maize pan-transcriptomes were obtained. The pan-transcriptome containing the coding part has, 33023 protein-coding genes. The pan-transcriptome containing the non-coding part has, 2215493 lncRNAs. For both pan-transcriptomes, annotation was performed, orthologous groups were determined, and the conserved and variable parts were characterized. Also, a search for novel lncRNAs was performed

separately for the pan-transcriptome with the non-coding part by aligning the obtained transcripts to databases of known lncRNAs and analyzing the expression of the identified novel lncRNAs. For the pan-transcriptome with a coding part, GO terms were analyzed and the expression of the identified new protein-coding genes was analyzed.
Conclusion: Two maize pan-transcriptomes containing a coding (protein coding genes) part and a non-coding (lncRNA) part were obtained. For both pan-transcriptomes, orthologous groups were identified, and both conserved and variable parts were identified. In addition, for the pan-transcriptome that includes non-coding regions, new lncRNAs were searched for.
Funding: This work was funded by the Kurchatov Genomic Center of the Institute of Cytology and Genotics of the Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation No. 075-15-2019-1662. The calculations were performed using the resources of the Bioinformatics Center.

Список литературы/References

1. Hirsch C.N. et al. Insights into the maize pan-genome and pan-transcriptome. *Plant Cell*. 2014;26(1):121-135

doi 10.18699/bgrs2024-1.4-14

Исследование особенностей профилей экспрессии РНК у пациентов с COVID-19

Шиманский В.С.[1, 2]*, Попов О.С.[1, 2], Клочкова Т.Г.[1], Апалько С.В.[1, 2], Сушенцева Н.Н.[1, 2], Асиновская А.Ю.[1], Мосенко С.В.[1, 2], Сарана А.М.[2], Щербак С.Г.[1, 2]

[1] *Санкт-Петербургское ГБУЗ «Городская больница № 40 Курортного района» отдела здравоохранения администрации Курортного района г. Санкт-Петербурга, Сестрорецк, Россия*

[2] *Санкт-Петербургский государственный университет Правительства Российской Федерации, Санкт-Петербург, Россия*

** shimansky.valya@yandex.ru*

Ключевые слова: дифференциальная экспрессия генов; COVID-19; секвенирование РНК

Мотивация и цель: Исследования, посвященные анализу периферических транскриптомных сигнатур COVID-19, активно ведутся во всем мире. Одно из таких исследований, проведенное китайскими учеными, показало 268 генов с измененной экспрессией, что привело к выявлению ключевых регуляторов убиквитинирования белков, включая 6 факторов транскрипции и 2 микроРНК [1]. Другое исследование выявило значительные различия в экспрессии 25 482 мРНК, 23 микроРНК и 410 длинных некодирующих РНК между пациентами с COVID-19 и контрольной группой. Отмечено, что изменения в экспрессии мРНК связаны с иммунным ответом, в частности с процессингом и представлением антигена, а также с активацией Т-клеток [2]. Транскриптомный анализ предоставляет возможность оценки молекулярного ответа организма на инфекцию и может помочь в выявлении потенциальных мишеней для терапии.

Целью данного исследования является дифференциальный анализ экспрессии генов для выявления набора регуляторных генов, влияющих на ключевые молекулярные и клеточные пути, вовлеченные в патогенез COVID-19.

Методы и алгоритмы: В исследовании приняли участие 146 пациентов с COVID-19. Все пациенты подписали информированное добровольное согласие на участие в исследовании. Исследование было одобрено экспертным советом по этике СПб ГБУЗ «Городская больница № 40» (протокол № 171 от 18 мая 2020 г.). Пациенты были разделены на основе степени тяжести заболевания: первая группа включала случаи с легким и умеренным течением, вторая группа – с тяжелым и крайне тяжелым. Дополнительно учитывался исход заболевания (выживание/летальный исход).

РНК была выделена из образцов крови вручную с помощью фенол-хлороформенного метода экстракции. Подготовка библиотек проведена с использованием реагентов KAPA RiboErase HMR (Roche), RNA Hyper Prep (Roche), KAPA Universal Adapter (Roche), KAPA UDI Primer Mixes (Roche). Конверсия проведена с использованием реагентов High-Throughput Sequencing Primer Kit (App-C) и MGIEasy Universal Library Conversion Kit (App-A) (MGI). Секвенирование полученных библиотек произведено на секвенаторе MGISEQ-2000 методом парно-концевого прочтения длиной 100 пар оснований на ячейке DNBSEQ-G400 High-throughput Sequencing Set (PE100, 360 Гб) (MGI).

Контроль качества прочтений выполнен в FastQC. Выравнивание прочтений на референсный геном версии hg38 выполнено программой STAR. Подсчет прочтений и аннотация выполнены программой featureCounts. Дифференциальная экспрессия генов оценивалась с помощью непараметрического критерия Манна–Уитни. Расчеты проводились в программном пакете R. Далее на базе платформы ge.genexplain.com произведен анализ обогащения дифференциально экспрессируемыми генами метаболических путей и поиск мастер-регуляторов.

Результаты: Мы провели анализ дифференциальной экспрессии генов, разделив образцы на группы по нескольким критериям. Уровень статистической значимости p-value ≤ 0.01.

1. Группы, выделенные по тяжести заболевания.

Мы выявили 806 генов с повышенной экспрессией и 3925 генов с пониженной экспрессией у группы с тяжелым течением COVID-19 по сравнению с группой с умеренным и легким течением. Анализ по базе данных Gene Ontology (GO) выявил следующие категории.

Гены с повышенной экспрессией концентрируются в метаболических путях, отвечающих за адаптивный иммунный ответ, сигналинг, межклеточную коммуникацию и ряд других. Гены с пониженной экспрессией сосредоточены в молекулярных путях, отвечающих за процессы метаболизма клеточных соединений азота и гетероциклов, генную экспрессию и т.д. (рис. 1).

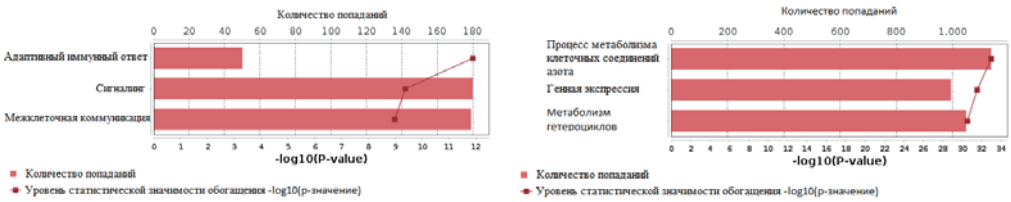


Рис. 1. Наиболее значимые результаты анализа обогащения молекулярных путей по GO сверхэкспрессированными генами (слева) и с пониженной экспрессией (справа)

Анализ мастер-регуляторов показал следующие результаты: *CDH2*, *CXCL9*, *KIT*, *IL13*, *EPHA4*, *FLT3*, *DUSP6*, *FGFR2*, *IL4* для генов с повышенной экспрессией; *RHOH*, *NFATC3*, *FXN*, *ADRB2*, *NTRK1*, *NAA30*, *ZC3H12D*, *MAF*, *KLRD1* для генов с пониженной экспрессией.

2. Группы, выделенные по исходу заболевания.

Из 9869 генов, прошедших по порогу статистической значимости, 348 показали повышенную экспрессию и 9521 пониженную у группы с летальным исходом по сравнению с группой выживших.

Проведя анализ обогащения по базе данных GO, мы выяснили, что гены с повышенной экспрессией чаще встречаются в метаболических путях, отвечающих за дегрануляцию нейтрофилов, нейтрофил-опосредованный иммунитет, ответ на внешние раздражители и ряд других. Гены с пониженной экспрессией сосредоточены в молекулярных путях, отвечающих за метаболизм клеточных соединений азота, соединений, содержащих азотистые основания, гетероциклов и т. д. (рис. 2).

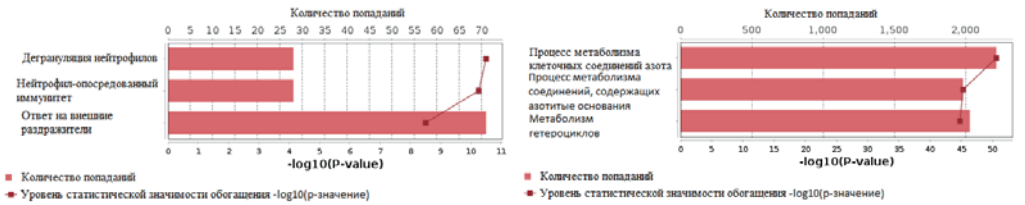


Рис. 2. Наиболее значимые результаты анализа обогащения молекулярных путей по GO сверхэкспрессированными генами (слева) и с пониженной экспрессией (справа)

Анализ мастер-регуляторов показал следующие результаты: *MAPK3*, *MMP9*, *PRL*, *MMP2*, *FLT3*, *S100AB*, *ITGB3*, *ITGA2B*, *PPARG* для генов с повышенной экспрессией; *PRNP*, *MAPK8*, *ERBB2*, *ZC3H122D*, *ITK*, *NAA30*, *TRIM3*, *ESR1*, *SH3RF2*, *ADRB2*, *KLRC1*, *KLRD1*, *PRKCH* для генов с пониженной экспрессией.

Выводы: По результатам данного исследования мы выявили ряд дифференциально экспрессируемых генов, потенциально объясняющих более тяжелое течение COVID-19, а также несколько мастер-регуляторов, чьи функции и роль в путях, обогащенных дифференциально экспрессированными генами, делают их потенциальными целями для биохимических и метаисследований. Наиболее интересными находками стали гены *ALOX15*, *IL4* и *KNDC1*. *ALOX15* имеет пониженную экспрессию в группах тяжелого течения и летального исхода, что свидетельствует об идущем в организме процессе воспаления и нарушенном механизме его терминации, потенциально ведущем к более тяжелому течению заболевания и исходу [3]. Также этот ген был выявлен как мастер-регулятор, что делает его кандидатом для исследования в качестве биомаркера и цели для терапии. У группы с летальным исходом снижена экспрессия гена *IL4*, кодирующего цитокин, являющегося ключевым регулятором гуморального и адаптивного иммунного ответа. Хотя прямой ассоциации *IL4* с тяжелым течением заболевания выявлено не было, этот ген представляет интерес для дальнейшего изучения [4]. Наш анализ показал, что *IL4*, предположительно, является одним из мастер-регуляторов для сверхэкспрессированных генов в группе с тяжелым течением заболевания. Ген *KNDC1*, участвующий в сигналинге, распознавании белков и функциональной регуляции, демонстрирует сниженную экспрессию у пациентов с тяжелым течением COVID-19 и летальным исходам.

Финансирование: Исследование выполнено в рамках проекта Санкт-Петербургского государственного университета (Pure ID 95412780) и при взаимодействии с базой Ресурсного центра Биобанк.

Expression profile analysis of COVID-19 patients

Shimansky V.S.[1, 2]*, Popov O.S.[1, 2], Klochkova T.G.[1], Apalko S.V.[1, 2], Sushentseva N.N.[1, 2], Asinovskaya A.Yu.[1], Mosenko S.V.[1, 2], Sarana A.M.[2], Shcherbak S.G.[1, 2]

[1] *City Hospital No. 40 of Health Department of the St. Petersburg Kurortniy District Administration, Sestroretsk, Russia*

[2] *St. Petersburg State University, Government of the Russian Federation, St. Petersburg, Russia*

* *shimansky.valya@yandex.ru*

Key words: differential expression; RNAseq; COVID-19

Motivation and Aim: Studies dedicated to the analysis of peripheral transcriptomic signatures of COVID-19 are actively being conducted worldwide. One such study by chinese scientists identified 268 genes with altered expression, leading to the discovery of key regulators of protein ubiquitination, including 6 transcription factors and 2 microRNAs [1]. Another study from China revealed significant differences in the expression of 25,482 mRNAs, 23 microRNAs, and 410 long non-coding RNAs between patients with COVID-19 and a control group. It was noted that changes in mRNA expression are associated with the immune response, particularly with antigen processing and presentation, as well as T-cell activation [2]. Transcriptomic analysis provides an opportunity to assess the molecular response of the body to infection and may assist in identification of potential targets for therapy.
The aim of this study is to perform a differential gene expression analysis to identify a set of regulatory genes that affect key molecular and cellular pathways involved in the pathogenesis of COVID-19.
Methods and Algorithms: The study involved 146 patients with COVID-19. All patients signed an informed voluntary consent to participate in the study. The research was approved by the ethics committee of the St. Petersburg State Budgetary Healthcare Institution "City Hospital No. 40" (protocol No. 171 of May 18, 2020). Patients were divided based on the severity of the disease: the first group included cases with mild and moderate progression, while the second group included severe and critically severe cases. Additionally, the outcome of the disease (survival/lethal outcome) was considered.
RNA was manually extracted from blood samples using the phenol-chloroform extraction method. Library preparation was conducted using KAPA RiboErase HMR (Roche), RNA Hyper Prep (Roche), KAPA Universal Adapter (Roche), and KAPA UDI Primer Mixes (Roche) reagents. Conversion was performed using High-Throughput Sequencing Primer Kit (App-C) and MGIEasy Universal Library Conversion Kit (App-A) (MGI) reagents. The sequencing of the obtained libraries was carried out on the MGISEQ-2000 sequencer by the paired-end reading method with a length of 100 base pairs on the DNBSEQ-G400 High-throughput Sequencing Set (PE100, 360 Gb) (MGI). Read quality control was performed in FastQC. Reads were aligned to the reference genome version hg38 using the STAR aligner. Read counts and annotation were performed with featureCounts. Differential gene expression was assessed using the Mann-Whitney nonparametric test. Calculations were conducted in the R software package. Subsequently, an analysis of the enrichment of differentially expressed genes in metabolic pathways and the search for master regulators were carried out on the ge.genexplain.com platform.
Results: We performed an analysis of differential gene expression, dividing the samples into groups based on several criteria. The level of statistical significance p-value is ≤ 0.01.
1. Groups divided by disease severity.
We identified 806 genes with increased expression and 3,925 genes with decreased expression in the group with severe COVID-19 compared to the group with moderate and mild progression. Gene Ontology (GO) enrichment analysis revealed the following categories. Genes with increased expression are concentrated in metabolic pathways responsible for adaptive immune response, signaling, intercellular communication, and others. Genes with decreased expression are focused on molecular pathways responsible

for the metabolism of cellular nitrogen compounds and heterocycles, gene expression, etc. (Fig. 1).

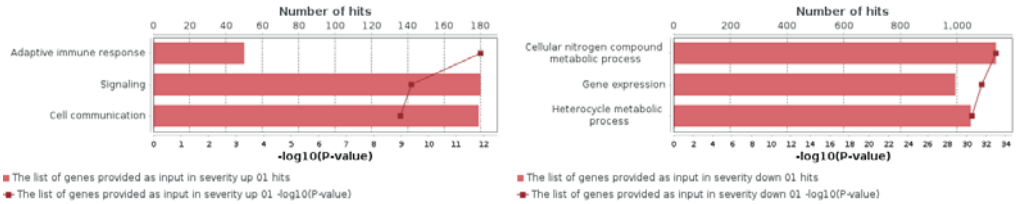


Fig. 1. The most significant results of the GO enrichment analysis of molecular pathways for overexpressed genes (on the left) and underexpressed genes (on the right)

The analysis of master regulators showed the following results: *CDH2*, *CXCL9*, *KIT*, *IL13*, *EPHA4*, *FLT3*, *DUSP6*, *FGFR2*, *IL4* for genes with increased expression; *RHOH*, *NFATC3*, *FXN*, *ADRB2*, *NTRK1*, *NAA30*, *ZC3H12D*, *MAF*, *KLRD1* for genes with decreased expression.

2. Groups divided by outcome of the disease.

Out of 9,869 genes that passed the threshold of statistical significance, 348 showed increased expression and 9,521 showed decreased expression in the group with lethal outcomes compared to the survivors' group. By conducting an enrichment analysis using the GO database, we found that genes with increased expression are more frequently involved in metabolic pathways responsible for neutrophil degranulation, neutrophil-mediated immunity, response to external stimuli, and others. Genes with decreased expression are concentrated in molecular pathways responsible for the metabolism of cellular nitrogen compounds, nucleobase-containing compound metabolic process, heterocycles, etc. (Fig. 2).

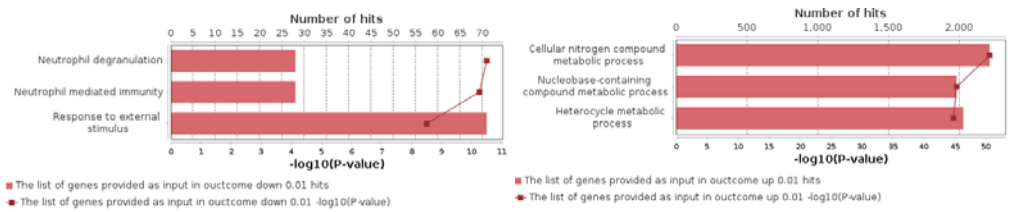


Fig. 2. The most significant results of the GO enrichment analysis of molecular pathways for overexpressed genes (on the left) and underexpressed genes (on the right)

The master regulator analysis revealed the following results: *MAPK3*, *MMP9*, *PRL*, *MMP2*, *FLT3*, *S100AB*, *ITGB3*, *ITGA2B*, *PPARG* for genes with increased expression; *PRNP*, *MAPK8*, *ERBB2*, *ZC3H122D*, *ITK*, *NAA30*, *TRIM3*, *ESR1*, *SH3RF2*, *ADRB2*, *KLRC1*, *KLRD1*, *PRKCH* for genes with decreased expression.

Conclusion: Based on the results of this study, we have identified a number of differentially expressed genes that potentially explain the more severe progression of COVID-19, as well as several master regulators whose functions and roles in pathways enriched with differentially expressed genes make them potential targets for biochemical and meta-research. The most interesting findings were the genes *ALOX15*, *IL4*, and *KNDC1*. *ALOX15* has decreased expression in groups with severe progression and lethal outcomes, indicating an ongoing inflammatory process in the body and a disrupted mechanism of its termination, potentially leading to a more severe course of the disease

and outcome [3]. This gene was also identified as a master regulator, making it a candidate for research as a biomarker and therapeutic target. The group with lethal outcomes showed decreased expression of the *IL4* gene, which encodes a cytokine that is a key regulator of humoral and adaptive immune responses. Although no direct association of *IL4* with severe disease progression was identified, this gene is of interest for further study [4]. Our analysis suggests that *IL4* is presumably one of the master regulators for the overexpressed genes in the group with severe disease progression. The *KNDC1* gene, which is involved in signaling, protein recognition, and functional regulation, shows decreased expression in patients with severe COVID-19 and lethal outcomes.

Funding: The study was supported by St. Petersburg State University, project ID: 95412780. The study was carried out in cooperation with the Core facility center Biobank.

Список литературы/References

1. Huang J. et al. Transcriptome analysis of peripheral blood mononuclear cells response in patients with severe COVID-19 reveals crucial genes regulating protein ubiquitination. *Med Sci Monitor*. 2022;28:e937532. doi 10.12659/MSM.937532
2. Li C. et al. Whole-transcriptome RNA sequencing reveals significant differentially expressed mRNAs, miRNAs, and lncRNAs and related regulating biological pathways in the peripheral blood of COVID-19 patients. *Mediators Inflammation*. 2021;2021:6635925. doi 10.1155/2021/6635925
3. Tian R. et al. ALOX15 as a suppressor of inflammation and cancer: lost in the link. *Prostaglandins Other Lipid Mediat*. 2017;132:77-83
4. Chang Y., Mengru B., Qinghai Y. Associations between serum interleukins (IL-1β, IL-2, IL-4, IL-6, IL-8, and IL-10) and disease severity of COVID-19: a systematic review and meta-analysis. *BioMed Res Int*. 2022;30:2022:2755246. doi 10.1155/2022/2755246

doi 10.18699/bgrs2024-1.4-15

Диссоциация ACME HS: универсальный метод фиксации-диссоциации клеток свежезамороженной ткани надпочечника человека для scRNA-seq

Щербакова А.[1]*, Уткина М.[1], Девятияров Р.[1, 2, 3, 4], Рябова А.[1], Урусова Л.[1]

[1] *ФГБУ «НМИЦ эндокринологии» Минздрава России, Москва, Россия*

[2] *Научный центр регуляторной геномики Института фундаментальной медицины и биологии Казанского федерального университета, Казань, Россия*

[3] *Высшая школа медицины, Университет Джунтендо, Токио, Япония*

[4] *Центр «Улучшение жизни технологиями будущего» (ЛИФТ), Москва, Россия*

* *nastya.shcherbakova1@gmail.com*

Ключевые слова: scRNA-seq; диссоциация тканей; свежезамороженные ткани; фиксированные клетки; надпочечники человека

Мотивация и цель: Главным этапом в секвенировании единичных клеток, определяющим биологическую релевантность и достоверность полученных scRNA-seq данных, является пробоподготовка образцов. Современные методы диссоциации не обеспечивают получение высококачественной суспензии живых одиночных клеток, особенно когда диссоциация затруднена морфологическими особенностями ткани или невозможна [1]. Мы представляем метод уксусно-метанольной диссоциации ACME HS (ACetic (acid)-MEthanol Hight Salt) для получения фиксированных и целых клеток из свежезамороженного материала [2]. В данной работе мы провели сравнительное исследование клеточной гетерогенности в наборах данных единичных клеток мозгового слоя надпочечников человека, полученных с помощью ACME HS и ферментативной диссоциации и изоляции ядер.

Методы и алгоритмы: С помощью метода scRNAseq мы получили данные транскриптомов от 15 образцов тканей медуллярного слоя надпочечника человека с общим количеством 94.807 клеток и ядер. Каунт матрицы были получены с помощью Cell Ranger (v6.1.1). Поиск дифференциально экспрессированных генов осуществлялся Seurat (v4.9.9 и v5.0.0). Интеграция выполнялась с использованием RunHarmony для обработанных Seurat объектов. Основные клеточные типы были аннотированы с помощью Conos43 (v1.5.0). Анализ скорости РНК выполнялся с помощью Velocyto (v0.17).

Результаты: Мы определили основные клеточные типы мозгового слоя надпочечника взрослого человека, включая медуллярные хромаффинные клетки, предшественники шванновских клеток (SCPs), васкулярные клетки кровеносных сосудов, перициты и иммунные клетки, которые были представлены во всех наборах данных, полученных с помощью ACME HS и ферментативной диссоциации, и отдельных ядер (рис. 1а). Хромаффинные клетки составляли большинство аннотированных клеточных типов (94 %) (рис. 1а). При изучении гетерогенности основных популяций хромаффинных клеток мы наблюдали потерю нескольких субкластеров в наборах данных ядер, за исключением ферментативных и ACME HS. С помощью UMAP визуализации было установлено, что хромаффинные клетки

экспрессировали основные тканеспецифичные маркерные гены, такие как *CHGA*, *SYP*, *DBH*, и *PNMT* [3, 4] для каждого из методов диссоциации (рис. 1b). При детальном анализе расширенной панели генов было установлено, что ACME HS сохраняет «типовые» гены, специфичные для хромаффинных клеток, включая *GSTA1*, *CYP11B1*, *HSD3B2*, *CYP17A1*, *TAC3*, *RNF17*, *MCOLN3*, *SIX3−AS1*, *CCN3*, *STAR* и *COX6*, экспрессия которых не наблюдалась в наборах данных для ферментативного и ядерного методов. Для того чтобы продемонстрировать согласованность между методом ACME HS и ферментативной диссоциацией, а также выявить их преимущества перед методом выделения ядер, нами была оценена динамика процесса транскрипции.

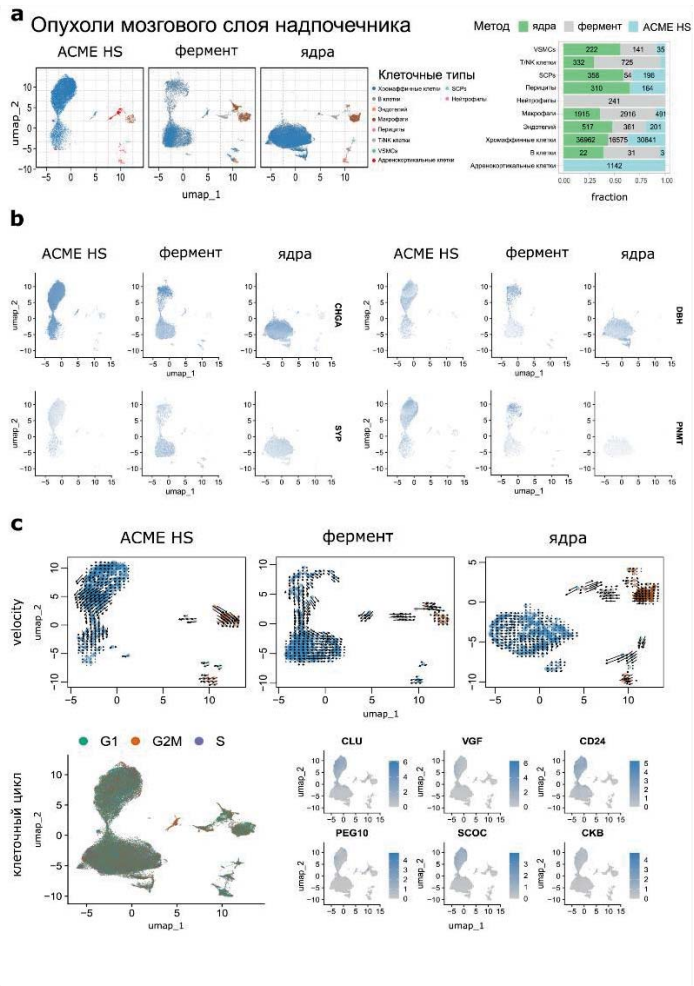


Рис. 1. Анализ данных scRNA-seq мозгового слоя надпочечника человека, полученных различными методами диссоциации – ACME HS, ферментативным и изоляцией ядер. **a.** Представленность клеточных типов и фракций среди трех методов диссоциации. **b.** UMAP визуализация экспрессии основных тканеспецифичных генов, включая *CHGA*, *SYP*, *DBH*, и *PNMT*. **c.** Анализ скорости РНК (Velocity) и клеточного цикла. Дифференциально экспрессируемые гены, связанные с контролем клеточного цикла, показаны на отдельных эмбеддингах

Анализ скорости РНК (velocity) обнаружил схожее движение дифференцирующихся хромаффинных клеток, включая движение к промежуточному состоянию дифференцировки и от него для двух методов диссоциации – ACME HS и ферментативного. Кроме того, мы идентифицировали активно пролиферирующие клетки, которые находились в начальной точке траектории скорости РНК и экспрессировали гены, ответственные за G2M/S фазовый переход – *CLU*, *VGF*, *CD24*, *PEG10*, *SCOC* и *CKB* (рис. 1c). Данные ядерного метода изоляции не содержали информации о направлении дифференцировки и связи между фазами клеточного цикла.

Выводы: В данном исследовании мы продемонстрировали эффективность и сопоставимость метода диссоциации ACME HS с ферментативным подходом, подчеркнув преимущества ACME HS перед изоляцией ядер. Транскрипционные профили ACME HS отражают тканеспецифичную экспрессию генов и подходят для изучения промежуточных состояний дифференцировки популяций раковых клеток и клеток-предшественников. ACME HS – надежный и универсальный метод получения фиксированных клеток из свежезамороженных тканей человека, разработанный для транскриптомики единичных клеток и обеспечивающий анализ с высоким разрешением.

Финансирование: Исследование поддержано Министерством науки и высшего образования Российской Федерации (договор № 075-15-2022-310 от 20 апреля 2022 г.).

ACME HS dissociation: an adaptable cell fixation-dissociation method for fresh-frozen human adrenal gland tissues for scRNA-seq

Shcherbakova A.[1]*, Utkina M.[1], Deviatiiarov R.[1, 2, 3, 4], Ryabova A.[1], Urusova L.[1]

[1] *Endocrinology Research Centre, Moscow, Russia*

[2] *Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[3] *Graduate School of Medicine, Juntendo University, Tokyo, Japan*

[4] *Life Improvement by Future Technologies (LIFT) Center, Moscow, Russia*

* *nastya.shcherbakova1@gmail.com*

Key words: scRNA-seq; tissue dissociation; fresh-frozen tissue; fixed single cell; human adrenal glands

Motivation and Aim: Sample preparation is a crucial step in single-cell RNA sequencing (scRNA-seq) of human tissues to obtain biologically relevant data. Current methods encounter challenges in obtaining high-quality single-cell suspensions, especially in tissues where live cell isolation is difficult [1]. To address this gap, we introduce ACME HS (ACetic (acid)-MEthanol Hight Salt), an optimized approach that facilitates the retrieval of intact and whole cells from fresh-frozen tissues [2]. In this study, we conducted a comparative analysis of cellular heterogeneity in datasets of single cells from the human adrenal medulla obtained from ACME HS, enzymatic dissociation, and nuclei isolation.

Methods and Algorithms: We analyzed transcriptomes from adult human adrenal glands using a single-cell transcriptomics approach (10X Genomics Chromium). Samples from

15 patients were included in the study for comparative analyses, totaling 94.807 cells and nuclei. The Raw sequenced reads were processed with 10X Cell Ranger (v6.1.1). Integrated embeddings and expression matrices were generated using Seurat (v4.9.9 and v5.0.0). Sample integration was carried out using RunHarmony on preprocessed Seurat objects. The major cell types were identified by the label propagation function in Conos43 (v1.5.0). Velocity analysis was performed with Velocyto (v0.17) and visualized using Velocyto.R (v0.6) on integrated embeddings.

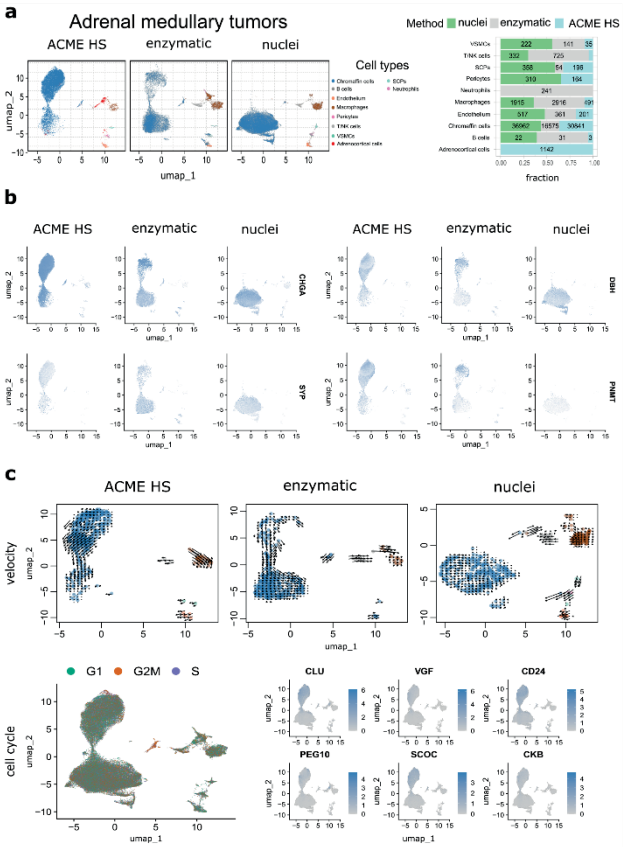


Fig. 1. scRNA-seq data analysis for human adrenal medullary glands by different dissociation methods – ACME HS, enzymatic, and nuclei isolation. **a.** Major cell type compositions and fractions among preparation methods. **b.** The key tissue-specific genes expression by UMAP visualization namely, *CHGA*, *SYP*, *DBH*, and *PNMT*. **c.** Velocity and cell cycle estimation for adrenal medullary tumor dataset. Examples of differentially expressed genes associated with cell cycle control are shown on the individual embeddings

Results: To compare the representation of distinct cell types among enzyme, nuclei, and ACME HS datasets, we generated integrated embeddings and annotations. Our analysis revealed all major cell types of the adult adrenal medulla, including numerous subtypes of medullary chromaffin cells, Schwann cell precursors (SCPs), vascular smooth muscle cells (VSMCs), pericytes, and immune cells represented and integrated in all three methods – enzymatic, ACME HS, and nuclei isolation (Fig. 1a). Chromaffin cells

constituted the majority of annotated cell types (94 %) (Fig. 1a). While studying the heterogeneity of major chromaffin cell populations, we observed the absence of several chromaffin subclusters in the nuclei datasets, which were consistently present in both ACME HS and enzymatic dissociation methods. We found that chromaffin cells expressed key tissue-specific marker genes such as *CHGA*, *SYP*, *DBH*, and *PNMT* in each method [3, 4] (Fig. 1b). To assess the broader efficacy of each method, we analyzed a comprehensive panel of genes specific for chromaffin cells and revealed that ACME HS retained the top genes, including *GSTA1*, *CYP11B1*, *HSD3B2*, *CYP17A1*, *TAC3*, *RNF17*, *MCOLN3*, *SIX3−AS1*, *CCN3*, *STAR*, and *COX6*, which were missing in the nuclei and enzyme datasets. In order to demonstrate consistency between ACME HS and enzymatic protocols together with their advantages over the nuclei isolation method we conducted RNA velocity analysis on individual samples. RNA velocity recapitulated the transcriptional dynamics within ACME HS and enzyme datasets, illustrating the overall movement of differentiating chromaffin cells, including movement towards and away from the intermediate differentiation state. Additionally, we identified cells associated with a proliferative state by determining their starting point in the velocity trajectory, marked by specific G2M/S phase genes such as *CLU*, *VGF*, *CD24*, *PEG10*, *SCOC*, and *CKB* (Fig. 1c). Nuclei-based data were largely lacking information regarding differentiation directions and connectivity between cell cycle phases.

Conclusion: In this study, we demonstrated the consistency of the ACME HS dissociation approach compared to the enzymatic method, highlighting its advantages over nuclei isolation. ACME HS transcriptional profiles maintain tissue-specific gene expression and are suitable for studying intermediate differentiation states of cancer and progenitor cell populations. Overall, ACME HS is a robust and practical method for obtaining fixed intact cells from primary human tissues, specifically designed for single-cell RNA transcriptomics and enabling high-resolution analysis.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2022-310 from April 20, 2022).

Список литературы/References

1. Machado L., Relaix F., Mourikis P. Stress relief: emerging methods to mitigate dissociation-induced artefacts. *Trends Cell Biol.* 2021;31(11):888-897
2. García-Castro H., Kenny N.J., Iglesias M. et al. ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. *Genome Biol.* 2021;22:89
3. Mete O. et al. Overview of the 2022 WHO Classification of Paragangliomas and Pheochromocytomas. *Endocr Pathol.* 2022;33:90-114
4. Konosu-Fukaya S. et al. Catecholamine-Synthesizing Enzymes in Pheochromocytoma and Extraadrenal Paraganglioma. *Endocr Pathol*. 2018;29:302-309

doi 10.18699/bgrs2024-1.4-16

Striatum transcriptome profile shift in repeated aggression mouse model quickly recovered to normal except for 12 circadian clock genes after fighting deprivation

Babenko V.[1], Redina O.[1], Smagin D.[1], Kovalenko I.[1], Galyamina A.[1], Babenko R.[1], Kudryavtseva N.[1, 2]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Pavlov Institute of Physiology, Russian Academy of Sciences, Saint Petersburg, Russia*

Key words: addiction; behavior; cAMP cascade; chronic social conflict; dopamine; dorsal striatum; endogenous opioids; nucleus accumbens; transcriptome

Both aggressive and aggression-deprived (AD) species represent pathologic cases intensely addressed in psychiatry and substance abuse disciplines. Previously, we reported that AD mice displayed a higher aggressive behavior score than the aggressive group, implying the manifestation of a withdrawal effect. We employed an animal model of chronic social conflicts, curated in our lab for more than 30 years. In the study, we pursued the task of evaluating key events in the dorsal striatum transcriptome of aggression experienced mice and AD species compared to controls using RNA-Seq profiling. Aggressive species were subjected to repeated social conflict encounters (fights) with regular positive (winners) experience in the course of 20 consecutive days (A20 group). This led to a profoundly shifted transcriptome expression profile relative to the control group, outlined by more than 1000 differentially expressed genes (DEGs). RNA-Seq cluster analysis revealed that elevated cyclic AMP (cAMP) signaling cascade and associated genes comprising 170 differentially expressed genes (DEGs) in aggressive (A20) species were accompanied by a downturn in the majority of other metabolic/signaling gene networks (839 DEGs) via the activation of transcriptional repressor DEGs. Fourteen days of a consecutive fighting deprivation period (AD group) featured the basic restoration of the normal (control) transcriptome expression profile yielding only 62 DEGs against the control. Notably, we observed a network of 12 coordinated DEG Transcription Factor (TF) activators from 62 DEGs in total that were distinctly altered in AD compared to control group, underlining the distinct transcription programs featuring AD group, partly retained from the aggressive encounters and not restored to normal in 14 days. We found circadian clock TFs among them, reported previously as a withdrawal effect factor. We conclude that the aggressive phenotype selection with positive reward effect (winning) manifests an addiction model featuring a distinct opioid-related withdrawal effect in AD group. Along with reporting profound transcriptome alteration in A20 group and gaining some insight on its specifics, we outline specific TF activator gene networks associated with transcriptional repression in affected species compared to controls, outlining *Nr1d1* as a primary candidate, thus offering putative therapeutic targets in opioid-induced withdrawal treatment.

doi 10.18699/bgrs2024-1.4-17

Approaches for sRNA analysis of human rna-seq data: comparison, benchmarking

Bezuglov V.[1, 2]*, Skakov I.[2], Shtratnikova V.[1], Sergeyev O.[1]**, Stupnikov A.[3, 4]

[1] *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

[2] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

[3] *Moscow Institute of Physics and Technology, Moscow, Russia*

[4] *National Medical Research Center for Endocrinology, Moscow, Russia*

** vitaly1530@gmail.com; ** olegsergeyev1@yandex.ru*

Key words: transcript annotation; Differential Gene Expression; small RNA; microRNA; piRNA; small RNA fragments; transcriptomics; benchmarking

Motivation and Aim: Transcriptomics analysis of various small RNA (sRNA) biotypes is a new and rapidly developing field. However, the bioinformatic analysis of NGS data for sRNA is prone to many challenges and not yet well established. Here, we attempt to identify the optimal pipeline configurations for each step of the sRNA analysis of human data, including read trimming, filtering, mapping, transcript abundance quantification, and differential expression (DE) analysis. Also we calculated the quality of obtained DE analysis signatures to estimate robustness of the obtained gene signature with robust and efficient rank statistics based approach.

Methods and Algorithms: Small RNA raw sequencing data from 7 published human studies [1–7]. All read adapters were removed, following lab kit protocols. To adjust the pipeline we used several trimming options for the upper bound. The trimmed reads were processed with various alignment (bowtie, hisat2, STAR, rsem) and pseudoalignment (salmon, kallisto) methods. Then three filtering strategies (by mean, median count and default DE analysis methods filtering) with several thresholds were applied to reduce noise. DE analysis was conducted using DESeq2, edgeR, limma, EBSeq and NOISeq packages. The number of DE transcripts cannot be used as a robust metric for DE analysis quality, as this approach does not account for false-positive results. To evaluate the expression signature quality, we applied the previously published Hobotnica approach [8]. Also we conducted an analysis with permuted group labels to detect false positive signatures and calculate FPR and Jaccard index.

Results: The most popular and usable methods of sRNA analysis were carried out. Based on our analysis, we suggest a pipeline that produces robust DE analysis results for sRNA transcripts, at least for categorical factors and two-group comparisons of biosamples (Table 1). We used existing tools to construct an optimal pipeline for quality sequencing data analysis, regardless of the differences in input data. To account for data variation in the original datasets, flexible trimming thresholds were applied. We suggest lower bound and upper bound of read length for trimming, thresholds for filtering expression data for the analysis of two groups. According to the results of DE analyses we suggest the DGE method for data with strong and well-detected signals.

Conclusion: The effect of various factors that impact the expression analysis of human sRNA at different stages of data processing were investigated. The optimal pipeline (trimming, aligning, assigning, filtering, DE analysis) parameters were suggested on Table 1, and an optimized pipeline for setting and running sRNA expression analysis

was proposed. Assessing the resulting expression signatures with rank-statistics-based inference suggests a way to estimate the quality of resulting signatures and performance of bioinformatic analysis for particular biological data. Results of this work were published [9] and new approaches (filtering thresholds and DE methods) were added to improve our suggestions.

Table 1. Recommended pipeline for sRNA analysis

Stage	Pipeline Command	Justification
Trimming	lower bound: 15 upper bound: Read length − 40% of adapter length	Retain sufficient and the same number of reads after trimming for downstream analyses for all datasets
Aligning	bowtie aligner with 1 mismatch allowed	The high alignment rate and H-score for all datasets
Assigning	ITAS	Optimized annotation for small RNA
Filtering	default DE method filter	Sufficient number of transcripts for the downstream analysis and higher H-score
DE analysis	DESeq2	Sufficient number of significant transcripts and high H-score

Funding: The study was supported by the Russian Science Foundation grant number 18-15-00202, https://rscf.ru/project/18-15-00202.

References

1. Donkin I., Versteyhe S., Ingerslev L.R. et al. Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. *Cell Metab.* 2016;23(2):369-378
2. Ingerslev L.R., Donkin I., Fabre O. et al Endurance training remodels sperm-borne small RNA expression and methylation at neurological gene hotspots. *Clin Epigenet.* 2018;10:12. doi 10.1186/s13148-018-0446-7
3. Hua M., Liu W., Chen Y. et al. Identification of small non-coding RNAs as sperm quality biomarkers for in vitro fertilization. *Cell Discovery.* 2019;5:20. doi 10.1038/s41421-019-0087-9
4. Wong R.K., MacMahon M. et al. A comparison of RNA extraction and sequencing protocols for detection of small RNAs in plasma. *BMC Genomics*. 2019;20(1):446
5. Huang G., Cao M. et al. Small RNA-sequencing identified the potential roles of neuron differentiation and MAPK signaling pathway in dilated cardiomyopathy. *Biomed Pharmacother*. 2019;114:108826
6. Kanth P., Hazel M.W. et al. Small RNA sequencing of sessile serrated polyps identifies microRNA profile associated with colon cancer. *Genes Chromosom Cancer*. 2019;58:23-33
7. Morgan C.P., Shetty A.C. et al. Repeated sampling facilitates within- and between-subject modeling of the human sperm transcriptome to identify dynamic and stress-responsive sncRNAs. *Sci Rep*. 2020;10:17498
8. Stupnikov A., Sizykh A., Budkina A. et al. Hobotnica: exploring molecular signature quality. *F1000Res*. 2021;10:1260. doi 10.12688/f1000research.74846.2
9. Bezuglov V., Stupnikov A., Skakov I. et al. Approaches for sRNA analysis of human RNA-seq data: comparison, benchmarking. *Int J Mol Sci.* 2023;24:4195. doi 10.3390/ijms24044195

doi 10.18699/bgrs2024-1.4-18

Alternative splicing in response to fungal niche shift

Chuan Xu*

Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China

** chuanx@sjtu.edu.cn*

Key words: Metarhizium robertsii; alternative splicing; niche shift; infection process

Motivation and Aim: Alternative splicing (AS) is a ubiquitous regulatory mechanism of gene expression in eukaryotes since it enables different mRNA and/or protein isoforms to be generated from the same gene [1]. Although AS plays important roles in fungi, how it changes along with fungal niche shift is understudied, especially for the pathogenic fungi [2].

Results: Taking entomopathogenic fungus Metarhizium robertsii as an example, we comprehensively investigated the genome-wide changes of AS in response to the niche shift during fungal infecting host. We analyzed 37 RNA-seq datasets of samples collected from three typical niches, including the saprophytic, infecting, and parasitic stages. Thousands of novel splicing events were then identified, but they were generally lowly occurred in comparison with those annotated in the genome. We found the infecting and parasitic stages gave rise to more AS events than the saprophytic stage. We also found that AS could act independently from the expression level through changing the structure of major isoforms and/or the co-splicing patterns. Eventually, a set of AS genes were uncovered as likely to be important in the niche shift during infection.

Conclusion: Together, our study reveals the significant changes and potential roles of AS in mediating pathogenic fungal adaptions to different niches in the process of infection.

References

1. Keren H., Lev-Maor G., Ast G. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet*. 2010;11(5):345-355. doi 10.1038/nrg2776
2. Fang S. et al. The occurrence and function of alternative splicing in fungi. *Fungal Boil Rev.* 2020;34(4):178-188. doi 10.1016/j.fbr.2020.10.001

doi 10.18699/bgrs2024-1.4-19

Exploring tissue-specific effects of 20-hydroxyecdysone through e23 hormone transporter expression in *Drosophila*

Evdokimova A.A.*, Vorobyeva N.E.

Ministry of Science and Higher Education of the Russian Federation, Institute of Gene Biology, RAS, Moscow, Russia

** sashkaa.eevd@gmail.com*

Key words: transcription; drosophila; enhancer; promotor; ecdysone

Motivation and Aim: The main steroid hormone in *Drosophila* is ecdysone, with its active form 20-hydroxyecdysone (20H). Elevated level of 20H facilitates the progression of several crucial phases in fly development. Double peaks of 20H during metamorphosis control the transition of larva to prepupa and then to pupa, activating fundamentally different transcriptional programs in fly tissues [1–3]. While larval tissues are targeted for destruction through apoptosis or autophagy, adult tissues begin to proliferate from the imaginal discs, moving on to differentiation [4, 5]. Impressively, all these programs are triggered by 20H. However, the knowledge about the tissue-specific action of steroid hormone is insufficient.

The aim of our work was to describe the mechanism of 20H-dependent transcriptional activation by studying *Drosophila* tissues.

Methods and Algorithms: To describe the mechanism of 20H-dependent activation of transcription in tissues we used a natural element of the ecdysone cascade – the E23 transporter. Previously it was shown that E23 can export 20H from cells [6, 7]. We made a transgenic line of *Drosophila melanogaster* hspE23-F26 that can express the ecdysone transporter E23 exogenously after heat shock treatment. We expressed E23 at a stage preceding the increase of hormone concentration during metamorphosis to suppress the response to ecdysone. We used RNA-seq and ChIP-seq data of tissues from experimental and control (not expressing E23) drosophila larvae to determine 20H target genes and to reveal the changes in transcriptional factors binding profiles of control tissues comparing to E23 expressing tissues.

Results: We identified ecdysone targets in salivary glands, which are consistent with the results of previous studies [8]. We also found that brain tissue is more resistant to ecdysone exposure at the studied stage as the previously described 20H targets were not expressed in brains of wandering larvae.

Furthermore, we described the changes that occur at the promoters and enhancers of 20H target genes when hormone concentration was lowered during the expression of the E23 transporter. We established that a larger group of promoters is regulated by 20H through stimulation of productive elongation. The decrease in the transcription of these promoters upon E23 expression is accompanied by a reduction in the level of the Pol II elongation complex in gene bodies. For the remaining smaller group of promoters increasing the 20H titer opens chromatin at the TSS and promotes Pol II recruitment. We concluded that 20H promotes the transcription of its primary targets in the salivary glands by stimulating Pol II recruitment or productive elongation.

Moreover, we showed that ecdysone stimulates transcription of target genes by creating new regions of open chromatin – enhancers, which complements previously obtained data on the mechanism of activation of ecdysone-dependent genes [9].
In addition, we show that salivary gland enhancers, organized when the 20H titer increases, are active in a tissue-specific manner.
Conclusion: In our work, we present an experimental system for investigating the role of steroid hormone in physiological transitions of *Drosophila melanogaster*. We perform multiple genome-wide profiling experiments to examine the role of ecdysone in gene regulation. Thus, we have demonstrated the use of the E23 transporter as an effective tool for studying the genome's response to ecdysone and, using this tool, described new mechanisms of gene transcription activation by the steroid hormone in drosophila.
Funding: The study is supported by Russian science foundation (No. 23-14-00184).

References

1. Kozlova T., Thummel C.S. Steroid regulation of postembryonic development and reproduction in *Drosophila*. *Trends Endocrinol Metab*. 2000;11(7):276-280. doi 10.1016/s1043-2760(00)00282-4
2. Thummel C.S. Molecular mechanisms of developmental timing in *C. elegans* and *Drosophila*. *Dev Cell* 2001;1:453-465
3. Zhimulev I.F., Belyaeva E.S., Semeshin V.F., Koryakov D.E., Demakov S.A., Demakova O.V., Pokholkova G.V., Andreyeva E.N. Polytene chromosomes: 70 years of genetic research. *Int Rev Cytol.* 2004;241:203-275
4. Yin V.P., Thummel C.S. Mechanisms of steroid-triggered programmed cell death in *Drosophila*. *Semin Cell Dev Biol.* 2005;16:237-243
5. Beira J.V., Paro R. The legacy of Drosophila imaginal discs. *Chromosoma.* 2016;125:573-592
6. Hock T., Cottrill T., Keegan J., Garza D. The E23 early gene of Drosophila encodes an ecdysone-inducible ATP-binding cassette transporter capable of repressing ecdysone-mediated gene activation. *Proc Natl Acad Sci USA.* 2000;97:9519-9524
7. Itoh T.Q., Tanimura T., Matsumoto A. Membrane-bound transporter controls the circadian transcription of clock genes in *Drosophila*. *Genes Cells*. 2011;16:1159-1167
8. Stoiber M., Celniker S., Cherbas L., Brown B., Cherbas P. Diverse Hormone Response Networks in 41 Independent *Drosophila* Cell Lines. *G3.* 2016;6(3):683-694
9. Uyehara C.M., Leatham-Jensen M., McKay D.J. Opportunistic binding of EcR to open chromatin drives tissue-specific developmental responses. *Proc Natl Acad Sci USA.* 2022;1199(40):e2208935119

doi 10.18699/bgrs2024-1.4-20

Overview of genes transcribed in the chicken lampbrush-stage oocyte and their key transcriptional regulators

Fedorov A.[1], Schelkunov M.[2, 3], Fedotova A.[2, 4], Krasikova A.[1]*

[1] *Saint-Petersburg State University, St. Petersburg, Russia*

[2] *Genomics Core Facility, Skolkovo Institute of Science and Technology, Moscow, Russia*

[3] *Institute for Information Transmission Problems, Moscow, Russia*

[4] *Lomonosov Moscow State University, Moscow, Russia*

* *alla.krasikova@gmail.com*

Key words: hypertranscription; lampbrush chromosomes; long non-coding RNA; maternal RNA; nuclear RNA-seq; oocyte nucleus; protein-coding genes; transcription factors; transcriptome profiling

Motivation and Aim: Chromosomes become lampbrushy due to an exceptionally high rate of nascent RNA synthesis. However, the set of genes transcribed on avian lampbrush chromosomes was not characterised. Recently, we performed nuclear and cytoplasmic transcriptome profiling of chicken oocytes at the lampbrush chromosome stage of oogenesis [1]. Here we have undertaken a functional characterisation of the genes expressed in chicken oocytes, tracing their expression levels during early embryogenesis and predicting their transcriptional regulators.

Methods and Algorithms: Gene set enrichment analysis to find over-represented gene ontology terms, key transcription regulators and histone modifications was performed using ENRICHR software. Tissue-specificity index (tau) was obtained from [2] and was used as a measure tissue-specificity of gene expression. K-means clustering and heatmap plotting were performed using Phantasus software. Visualisations were done with R software, version 4.2.2.

Results: A total of 77 and 95 % of all reads from the chicken oocyte nuclear and cytoplasmic total RNA libraries, respectively, were mapped within annotated genes. Transcripts for 10697 protein-coding genes and 2488 long non-coding RNA genes in the oocyte nucleus and 11402 protein-coding genes and 910 lncRNA genes in the oocyte cytoplasm were detected. The oocyte nucleus contains nascent transcripts (characterised by the presence of reads corresponding to introns) for more than 85 % of the protein-coding genes whose transcripts are found in the oocyte cytoplasm. In order to perform a functional characterisation of genes expressed in the oocytes, we traced their expression level during early embryogenesis, identified seven clusters of co-expressed genes and found associated biological processes. For that, chicken oocyte transcriptome data was combined with chicken embryos expression dataset [3]. Next we compared RNA repertoire in the oocyte nucleus and cytoplasm with the transcription profiles in different chicken organs. The identified co-expression clusters can be divided into 4 groups – those where the maximum expression is in oocytes, or in early embryos before zygotic genome activation, or in the embryos after genome activation, or in differentiated organs. The data suggest that the set of genes transcribed in lampbrush stage oocytes is enriched in genes with a broad expression pattern (housekeeping genes). Genes for lncRNAs were enriched in two clusters – the cluster of nuclear retained RNAs, suggesting their function in intranuclear events, and the cluster of genes not transcribed in the oocyte, consistent with their role as tissue-specific regulators activated at later stages of development.

Enrichment analysis was used to find key transcription regulators from ChEA 2022 database, which are associated with subsets of genes with the highest and the lowest levels in chicken oocyte. Hypertranscription on lampbrush chromosomes can be activated by predicted transcriptional regulators MYC, YY1, E2F1, KDM5B, ZFX, and others. At the same time, repression of the inactived genes in avian lampbrush chromosomes can be controlled by epigenetic mechanisms involving Polycomb complexes PRC1 (RNF2, BMI1, CBX2, PHC1) and PRC2 (SUZ12, JARID2, EZH2, EED, MTF2).

Conclusion: Strong overlap between sets of genes whose transcripts were found in the oocyte nucleus and cytoplasm indicates transcription within the oocyte nucleus at the lampbrush stage. The total number of genes transcribed in the oocyte itself is comparable to that in other tissues. Genes expressed in the oocyte nucleus are characterised by a broader expression profile compared to genes whose transcripts were absent in the oocyte. Clusters of genes whose high transcript levels correlate with each other in the oocyte nucleus and cytoplasm are mainly associated with essential cellular processes including transcription, mRNA catabolism, piRNA metabolism, DNA repair, RNA splicing, protein synthesis, DNA replication, cell cycle progression and mitochondria functions. We assume that maternally deposited transcripts derived from the oocyte nucleus are required for the cellular processes associated with oogenesis and early embryogenesis.

Funding: The research was supported by the Russian Science Foundation (grant #19-74-20075) and was performed using the equipment of the Genomics Core Facility (Skoltech) and Resource Center “Molecular and Cell Technologies” (Saint-Petersburg State University).

References

1. Krasikova A. et al. The first chicken oocyte nucleus whole transcriptomic profile defines the spectrum of maternal mRNA and non-coding RNA genes transcribed by the lampbrush chromosomes. *bioRxiv*. 2024. doi 10.1101/2024.02.05.577752
2. Bush S.J. et al. Combination of novel and public RNA-seq datasets to generate an mRNA expression atlas for the domestic chicken. *BMC Genomics*. 2018;19:594. doi 10.1186/s12864-018-4972-7
3. Hwang Y.S. et al. The first whole transcriptomic exploration of pre-oviposited early chicken embryos using single and bulked embryonic RNA-sequencing. *GigaScience*. 2018;7:1-9. doi 10.1093/gigascience/giy030

doi 10.18699/bgrs2024-1.4-21

Roles of the transcriptional R-loop forming sequences (RLFS) and RLFS(+) R-loops in normal and cancer cell states. Critical review

Kapali O., Adala J., Kuznetsov V.A.

Department of Urology, Department of Biochemistry and Molecular Biology, SUNY Upstate Medical University, Syracuse, USA

Key words: R-loops; R-loop forming sequence; transcription regulation; cancer; regulatory sites

Motivation and Aim: Transcriptional R-loops are dynamical three-stranded guanine-rich RNA-DNA hybrid structures in which a nascent RNA is hybridized to a template DNA strand, with the other (non-template) DNA strand looped out. We have proposed that such single-stranded DNA sequences determine the initiation and stabilization of R-loops genome-wide and found that they are ubiquitous in eukaryotic genomes [1–4]. We called a family of these sequences the R-loop forming sequences (RLFS).

We have introduced the Quantitative models of R-Loop Forming Sequences (QmRLFS) tools [1, 2, 4]. As per QmRLFS model, RLFS is a non-template guanine-rich sequence consisting of three zones: R-loop initiation zone (RIZ, includes few G repeats), linker sequence, and R-loop elongation zone (REZ, high rich G sequence). QmRLFS identifies the sizes, *positions, and boundaries* of RLFS with high specificity and sensitivity. The model predicts the size and boundaries of the R-loop at an accuracy of 86–92 % in vitro [1, 2, 4, 5] providing objective controls and improved R-loop detection methods [6, 7]. We have discovered that in the human genome RLFSs are present in the vicinity of transcription regulatory sites (Transcription Start Site (TSS), Transcription Termination Site (TTS), Intron-Exon junctions, transcribed enhancers, etc.) of more than 75 % of protein-coding genes, about 25 % of non-coding transcribed loci and they are highly enriched with G-quadruplexes [3, 4]. We have classified RLFS regulatory functions genome-wide in humans and other eukaryotes [3–5]. About 90 % of the predicted RLFS(+) merged regions formed R-loops/RNA:DNA hybrids were supported by genome-wide data [2–4]. Our model and facts have been highlighted in recent reviews and used in databases and web tools [7–9]. Over 1,000 proteins have been reported to interact with RNA:DNA hybrids/R-loops, most of which are RLFS-positive, suggesting the existence of a robust RNA:DNA hybrid-protein interactome [10, 11]. These findings have suggested multiple structural and functional roles of the R-loops/RNA:DNA hybrids in cell physiology, stress response/adaptation, and pathobiology, including cancer [12]. In vitro and cell line model studies have shown that R-loops/RNA:DNAs can provide essential roles in genotoxic stress, genome instability, transcription alterations, pro-oncogenic mutagenesis, and links with several other cancer hallmarks. It has been proposed that R-loops may be relevant to oncogenesis and aggressive cancer phenotypes [1–4, 12]. Despite recent discoveries in R-loop biology, our current knowledge of the effects of R-loops and RLFS in the formation and stability of RNA:DNA hybrids/R-loops in cancer is poorly understood. Lack of

knowledge about R-loop initiation, formation, and regulation in cancer biology and limitations of detection methods prevent clinical studies and evaluation of R-loops and their subsets' roles in cancer.

The objective of our review is to consider the computationally predicted RLFSs and R-loop biology studies in the context of improvement of quality control, accuracy, reproducibility, and resolution experimental methods. We also consider the perspective of integrative computational biology, bioinformatics and functional genomics, single-cell and image analyses of R-loops, and mathematical modeling to move forward to cancer R-loop pathobiology and clinical oncology needs.

Consideration: Perspectives of RLFS, RNA-DNA hybrids and R-loop detection genome-wide technologies, big data collection and analyses. R-loop detection technologies, reproducibility, standards, quality control, specificity, and sensitivity, analytical tools and computational methods, and databases. Development of biological models, and the integrative databases of comprehensive clinical data study. Roles of RLFS-defined RNA:DNA hybrids in the pathogenesis, cells and clinical samples heterogeneity, diagnostic and prognosis of cancer using breast cancer data analysis as our analysis example.

Conclusion: The RLFS subset due to their sequence uniqueness and commonness in transcription regulatory sites, as well as the high accuracy of the computationally predicted RLFS and corresponding R-loop positions, may be defined as novel structural and functional determinants that are essential to cell physiology and genetic disease (including cancer). After adopting them for clinical practice and their detection, such structures may be useful targets in cancer diagnostics, personalized treatment, and predictors.

References

1. Wongsurawat T. et al. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res.* 2012;40(2):e16
2. Jenjaroenpun P. et al. QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. *Nucleic Acids Res*. 2015;43(W1):W527-W534
3. Jenjaroenpun P. et al. R-loopDB: a database for R-loop forming sequences (RLFS) and R-loops. *Nucleic Acids Res*. 2017;45(D1):D119-D127
4. Kuznetsov V.A. et al. Toward predictive R-loop computational biology: genome-scale prediction of R-loops reveals their association with complex promoter structures, G-quadruplexes and transcriptionally active enhancers. *Nucleic Acids Res*. 2018;46(15):7566-7585
5. Yeo A.J. et al. R-loops in proliferating cells but not in the brain: implications for AOA2 and other autosomal recessive ataxias. *PLoS One*. 2014;9(3):e90219
6. Safari M. et al. R-loop-mediated ssDNA breaks accumulate following short-term exposure to the HDAC inhibitor romidepsin. *Mol Cancer Res*. 2021;19(8):1361-1374
7. Lin R. et al. R-loopBase: a knowledgebase for genome-wide R-loop formation and regulation. *Nucleic Acids Res*. 2022;50(D1):D303-D315.
8. Miller H.E. et al. Exploration and analysis of R-loop mapping data with RLBase. *Nucleic Acids Res.* 2023;51(D1):D1129-D1137
9. Shi X., Teng H., Sun Z. An updated overview of experimental and computational approaches to identify non-canonical DNA/RNA structures with emphasis on G-quadruplexes and R-loops. *Brief Bioinform*. 2022;23(6):bbac441. doi 10.1093/bib/bbac441
10. Wang I.X. et al. Human proteins that interact with RNA/DNA hybrids. *Genome Res.* 2018;28(9):1405-1414
11. Cristini A. et al., RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. *Cell Rep*. 2018;23(6):1891-1905
12. Gaillard H., García-Muse T., Aguilera A. Replication stress and cancer. *Nat Rev Cancer*. 2015;15(5):276-89. doi 10.1038/nrc3916

doi 10.18699/bgrs2024-1.4-22

DEAr – tool for differential expression gene analysis

Karpenko D.*
National Medical Research Center for Hematology, Moscow, Russia
** d_@list.ru*

Key words: DEAr; differential expression; housekeeping genes; statweight; recursive calculations; ranged values; missing values; normalization

Motivation and Aim: Differential gene expression is a standard way to study and to describe cells and tissues. Directed studies or clinical tests often use real-time PCR with limited number of measured genes. For individual samples, we can't be sure about quantity and quality of analyzed RNA. An important stage in the proceeding is a normalization. Normalization methods can be generalized to the idea of establishing a baseline from a subset of stable genes, and then representing gene expressions as differences from this baseline. If we compare different cells under different circumstances, some genes are considered to establish a baseline could be unstable in particular samples or conditions. In case of collecting clinical data for years or from multiple centers, datasets could include different subsets for normalization, forming that way a mosaic data. This leads to natural difficulties in data processing. There are no automated solutions or the strict protocol, especially for inconsistent real-time PCR data. Therefore, I decided to create an automated differential expression calculation tool that could handle mosaic data and be resistant to individual outliers.
Methods and Algorithms: Instead of selecting preferable for normalization genes, the original algorithm identifies common expression pattern for genes in analysis. Pattern identification is based on expanded data of pairwise comparisons between genes of a sample. It negates system shift in data from RNA quantity, but increases data complexity. Through several cycles of adjusting and convolution, cohesive data for the whole dataset is formed. In each cycle, the algorithm identifies how to adjust each expression value in samples to reach a common pattern. As the algorithm changes all values simultaneously, recursive calculations are required to reach a stable result. Due to multiple outliers in data, high robustness is required from the algorithm. I used statweight to make the necessary distinction between noise and reproducible patterns. A statweight is a measure of how reproducible a particular element of the data, and is interpreted as an effective number of measurements. Applied statweights are formed as an exponential penalty for deviations from the common pattern. An additional penalty is applied according to the inaccuracy of a measurement, implemented in such a way that each value is considered as a range. This allows both accurate and inaccurate values to be used in the same calculation, while distinguishing their influence. While the statistical weights are based on the reproducibility of the data, I have also applied predictive weights based on external knowledge of the stability and instability of genes. This allows the algorithm to operate on a combination of information from external and internal data. This means that the algorithm can work on unknown gene profiles, but additional information can be used to improve performance. Alike outliers and uncertain values, missing values don't contribute to a common pattern and that way provide no obstacles for analysis. This follow general theme of making tools for problematic data with ranged

and missing values [1]. In summary, the DEAr (Differential Expression Analyzer) tool uses a highly recursive algorithm developed in Python v.3.7, packaged in an executable file. I presented algorithm on previous BGRS2022 [2]. The extended algorithm presentation is available in a preprint [3].

Results: I present ready to use software available within the preprint version of the study [3]. Analysis of the real data from another study [4] showed that genes considered as stable were unstable for certain tissues. On the data from the experiment with mosaic deletion, DEAr performed calculations as for initial data. Standard protocols don't provide a ready to use solution and require manual data preparation to compensate for missing data by inputting or discarding of whole gene for all samples. In the testing data, the single value for the gene of interest was manually shifted as it was contaminated with PCR product. DEAr recognized it as an outlier from three technical replicates due to its highly robust core algorithms using statweight, while standard geometric mean demonstrated shift in the resulting evaluation. DEAr allows aggregation of data or knowledge not only from same experiments but also from external sources, evaluated as how we are certain for a value in sample or for a gene expression stability in general. The program can be used as a single executable file that gets and returns data in Excel format. No additional programming skills are required.

Conclusion: I present the original tool DEAr for differential expression analysis and the original implementation of the robust computations. The tool utilizes information of different origin to provide a conclusion. The applied approach uses information of each measurement accuracy. DEAr works in the Microsoft Windows OS using data in Excel format. As no additional programming skills are required, DEAr will be a handy tool for classical biologists or medical scientists. The program is available to any independent researcher who wishes to use it for non-commercial purposes, provided that the name of the program and the article [3] are correctly cited. Use by non-academics requires a license.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Karpenko D.V., Bigildeev A.E. Small groups in multidimensional feature space: two examples of supervised two-group classification from biomedicine. *J Bioinform Comput Biol*. 2023;21(6):2350025. doi 10.1142/S0219720023500257
2. Karpenko D.V. Recursive matrix algorithm for calculating differential expressions. In: BGRS/SB-2022. Novosibirsk, 2022;1101-1102. doi 10.18699/SBB-2022-660
3. Karpenko D.V. DEAr – Differential expression analyzer. 2023. doi 10.21203/RS.3.RS-2957165/V3
4. Karpenko D., Dorofeeva A., Petinati N., Shipounova I., Drize N., Bigildeev A. Functional characteristics of the mouse Il1b promoter in various tissues before and after irradiation. *DNA Cell Biol*. 2020;39:790-800. doi 10.1089/DNA.2019.5310

doi 10.18699/bgrs2024-1.4-23

Antibacterial activity of rainbow trout plasma: *in vitro* assays and proteomic analysis

Mizaeva T.[1], Alieva K.[1], Zulkarneev E.[2], Pomazkin V.[3], Kurpe S.[4], Matrosova S.[5], Borvinskaya E.[3]*, Sukhovskaya I.[5, 6]

[1] *G.N. Gabrichevsky Research Institute for Epidemiology and Microbiology, Moscow, Russia*

[2] *Plague Control Center, Federal Service on Consumers' Rights Protection and Human Well-Being Surveillance, Moscow, Russia*

[3] *Institute of Biology, Irkutsk State University, Irkutsk, Russia*

[4] *Institute of Biochemistry after H. Buniatyan National Academy of Sciences of the Republic of Armenia, Yerevan, Armenia*

[5] *Institute of Biology, Ecology and Agricultural Technologies of the Petrozavodsk State University, Petrozavodsk, Russia*

[6] *Institute of Biology of the Karelian Research Centre, RAS, Petrozavodsk, Russia*

** borvinska@gmail.com*

Key words: rainbow trout; antimicrobial proteins; *Aeromonas hydrophila*; proteomics

Motivation and Aim: The ability of body fluids and secretions to kill bacteria is an important aspect of immune defence. Although many antimicrobial factors have already been discovered in fish plasma, the biochemical and physiological consequences of contact between bacteria and the complex of bacteria-binding proteins present in the blood remain poorly understood. Previously, a simple approach for screening potential immunomodulators was developed using live bacteria as bait for affinity-bound proteins from biological fluids [1–3]. The high-throughput mass spectrometry used in these studies provides unique information on protein–protein interactions between microorganisms and the host blood proteins as well as real-time monitoring of adaptive metabolic rearrangements of pathogens during invasion. This approach is therefore very promising for screening previously uncharacterised proteins with antimicrobial activity in biological fluids from different species.

In this study, we investigated the natural antibacterial activity of blood plasma of cultured rainbow trout *Oncorhynchus mykiss*. First, we studied the growth dynamics, morphology, and protein profile of the fish pathogen *Aeromonas hydrophila* when incubated in plasma from trout from two different fish farms. Second, a comparative analysis of fish plasma samples with different levels of antimicrobial activity was performed. To identify molecules with signaling functions and bactericidal activity, the study design also included an investigation of trout proteins interacting with live *A. hydrophila in vitro*. Finally, changes in the pathogen proteome in response to contact with the host blood plasma were studied to elucidate the mechanisms of action of fish antimicrobial factors.

Methods and Algorithms: Fish blood plasma was collected from two independent populations of rainbow trout *O. mykiss.* Fish from the first farm (Farm A, Lake Onego, Northwest Russia) were characterised as relatively healthy. Fish from the second fish farm (farm N, Angara River, Eastern Siberia) have chronic bacterial disease associated with intestinal inflammation, rectal prolapse, and skin hyperemia. Sequencing of the

hypervariable V3-V4 region of the 16S rRNA gene in inflamed tissue revealed an overwhelming number of reads belonging to the salmonid pathogen *F. psychrophilum*.
Native blood plasma from fish caught in farm A was pooled and further referred to as AP and plasma from fish caught in farm N was referred to as NP. The low molecular weight fraction (< 10 kDa) was obtained by plasma filtration using Amicon Ultra-0.5 Millipore concentrators. Denatured plasma was obtained by heating at 56 °C for 30 min.
The antimicrobial activity of native, filtered and denatured AP and NP trout plasma was assessed by the growth inhibition zone at the plasma application sites on Mueller–Hinton agar (MHA) seeded with a bacterial culture of *Aeromonas hydrophila*. The growth dynamics of *A. hydrophila* incubated in fish plasma was assessed by the optical density of the bacterial biomass at 620 nm. The viability of *A. hydrophila* after 4 h incubation in fish plasma was assessed by counting the subsequent growth of bacterial colonies on 1.5 % MHA. Morphology and cell integrity of *A. hydrophila* after incubation with fish plasma were assessed by fluorescence microscopy using the LIVE/DEAD® BacLight Bacterial Viability Kit (Invitrogen).
For the proteomic assay, the native plasma was analysed as is, while the bacterial cells were pre-treated for protein isolation. After 4 h incubation in AP, NP or denatured AP plasma, suspensions of *A. hydrophila* were washed in buffer to remove unbound trout proteins by a triple sedimentation-resuspension procedure, followed by cell lysis in 2.5 % SDS. The collected samples were homogenised and trypsinised according to standard procedures.
Proteomic analysis of trout and bacterial peptides was performed using an Ultimate 3000 RSLCnano chromatographic HPLC system (Thermo Scientific) coupled to a Q-exactive HFX mass spectrometer (Thermo Scientific). Protein identification in the mass spectra was performed using MaxQuant v. 1.6.3.4 via the Andromeda search algorithm. The UniProt sequence databases UP000193380 (*O. mykiss*) and UP000000756 (*A. hydrophila*) were used to identify proteins in the samples. Protein abundance analysis was based on label-free quantification (LFQ). Differential expression between samples was analysed using the *Limma* R package.
Results: The study demonstrated the inhibition of *A. hydrophila* growth by native blood plasma obtained from trout infected with *F. psychrophilum* (antimicrobial plasma, AP), suggesting the production of bacteriostatic humoral immune factors and/or antibodies in diseased fish. No antimicrobial activity against *A. hydrophila* was observed in the plasma from healthy trout (non-antimicrobial plasma, NP), demonstrating that antigen loading is necessary for the humoral immune response in cultured rainbow trout. The level of bactericidal activity of the 10-kDa cut-off filtered or denaturated AP and NP plasma was low, suggesting that the inhibitory effect of AP is related to the protein fraction.
After 4 hours of incubation in AP, but not NP plasma, *A. hydrophila* lost its ability to divide on the culture media, although, according to lifetime staining, most of the cells were not destroyed. Microscopy revealed changes in the morphology of AP-treated bacterial cells, which were irregular in length, chained together and agglutinated.
Circulating immunoglobulins have been suggested as candidates for the major antimicrobial factor in the AP plasma. Immunoglobulins isoforms were shown to differ in their ability to bind bacterial cells. The isoform abundant in AP plasma was found to adhere to *A. hydrophila* cells treated with both AP and NP. In turn, another two Ig isoforms were identified exclusively in the lysate of bacteria treated with bacteriostatic AP. These proteins were not detectable in the mass spectra of native plasma, suggesting

that minor but highly selective components of AP could play role in pathogen recognition.

Increased level of the circulating isoform of C-type lectin was found in AP plasma, but not in the washed lysate of either AP- or NP-treated bacteria. In turn, another C-type lectin (ladderlectin) was found to be retained exclusively in the lysate of *A. hydrophila* treated with antimicrobial AP plasma. Thus, lectins selectively retained in bacterial lysates can be suggested as another candidate for the role of the antimicrobial factor in trout challenged with an *F. psychrophilum* infection.

Finally, we have shown that treatment with AP plasma severely affects protein biosynthesis in bacteria, including the loss of components of the translation machinery. This led to the disorganisation of core cellular processes and to the observed inhibition of cell growth, although it is not yet clear which host molecules might be the trigger for this process.

Conclusion: In our screening study, we found that antimicrobial proteins in the plasma of cultured rainbow trout infected with *F. psychrophilum* inhibit the growth of *A. hydrophila in vitro* by arresting protein biosynthesis in the bacteria rather than by causing cell perforation. Cross-reactive antibodies and the antimicrobial protein ladderlectin, which is capable of cell agglutination, are thought to be responsible for the bacteriostatic properties of the fish plasma. The applied affinity approach, in which bacterial cells were used as bait, greatly expanded the range of plasma proteins analysed, and allowed the capture of minor components that selectively bind to antigens and are thus directly involved in immune defence.

Funding: The study is supported by the Russian Science Foundation (No. 20-66-47012, interdisciplinary project performed in association with the Irkutsk State University, Irkutsk, Russia).

References

1. Dong M., Liang Y., Ramalingam R., Tang S.W., Shen W., Ye R., Gopalakrishnan S., Au D.W.T., Lam Y.W. Proteomic characterization of the interactions between fish serum proteins and waterborne bacteria reveals the suppression of anti-oxidative defense as a serum-mediated antimicrobial mechanism. *Fish Shellfish Immunol.* 2017;62:96-106. doi 10.1016/j.fsi.2017.01.013
2. Liu Y., Zhang H., Liu Y., Li H., Peng X. Determination of the heterogeneous interactome between edwardsiella tarda and fish gills. *J Proteom.* 2012;75:1119-1128. doi 10.1016/j.jprot.2011.10.022
3. Li H., Zhu Q., Peng X., Peng B. Interactome of E. piscicida and grouper liver proteins reveals strategies of bacterial infection and host immune response. *Sci Rep.* 2017;7:39824. doi 10.1038/srep39824

doi 10.18699/bgrs2024-1.4-24

Ontology analysis of big transcriptomic data and differential gene expression

Oshchepkov D.[1], Chadaeva I.[1]*, Kozhemyakina R.[1], Zolotareva K.[1], Khandaev B.[1], Sharypova E.[1], Ponomarenko P.[1], Bogomolov A.[1], Klimova N.V.[1], Shikhevich S.[1], Redina O.[1], Kolosova N.G.[1], Nazarenko M.[2], Kolchanov N.A.[1], Markel A.[1], Ponomarenko M.[1]

[1] *Institute of Cytology and Genetics, Novosibirsk, Russia*

[2] *Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, Russia*

* *ichadaeva@bionet.nsc.ru*

Key words: hypertension; stress reactivity; molecular marker; age-related disease; differentially expressed gene; functional annotation of genes; gene network; correlation; principal component; bootstrap

Motivation and Aim: The aim of this study was to conduct an ontological analysis of the transcriptome of two lines of gray rats, tame and aggressive, as well as identified differentially expressed genes, and compare them with human transcriptomic data in accordance with the phenotype of a specific disease.

Methods and Algorithms: To analyze the DEGs, we used our previously obtained high-throughput RNA sequencing data from four brain regions of gray rats of the two strains: the hypothalamus [1], the hippocampus [2], the periaqueductal gray (PAG) [3], and the midbrain tegmentum (MTg) [4]. DEGs were subjected to principal component analysis using the PAST4.04 software package. We used available transcriptome data from tissues from patients with diseases compared with apparently healthy controls, model animals compared with controls, and domestic animals compared with wild animals. Using the PubMed database, we characterized rat DEGs and human DEGs in terms of the clinical implications of their under- or over-expression for disease. All these data were statistically significant according to Fisher's Z test, adjusted for multiple comparisons.

Results: The hippocampal transcriptome of tame and aggressive gray rats was analyzed using an ontological approach to investigate molecular markers of hypertension. It is known that increased stress reactivity increases the risk of developing arterial hypertension. In this regard, an analysis of the identified DEGs was carried out in the ontological data space defined by the concepts “Gene Expression”, “Stress Reactivity/Domestication” and “Hypertension”. This approach collected all transcriptomic data from animal models of hypertension available at the time of the study. These are 14 models for 4 animal species in which hypertension was studied: rats, mice, rabbits and chickens, for a total of 4216 DEGs. We also collected all available transcriptomic data from studies of patients with hypertension compared with apparently healthy controls, for which a total of 7865 DEGs were identified (Figure 1). In total, these two samples amounted to 12081 DEGs. Thus, an analysis that started with 42 DEGs resulted in the need to consider a big data sample that was 300 times larger in size. Analysis in each of the samples revealed a significant correlation between changes in expression levels obtained in experiments with stress reactivity in tame and aggressive

rats, as well as in experiments with hypertension for homologous genes in model animals and humans.

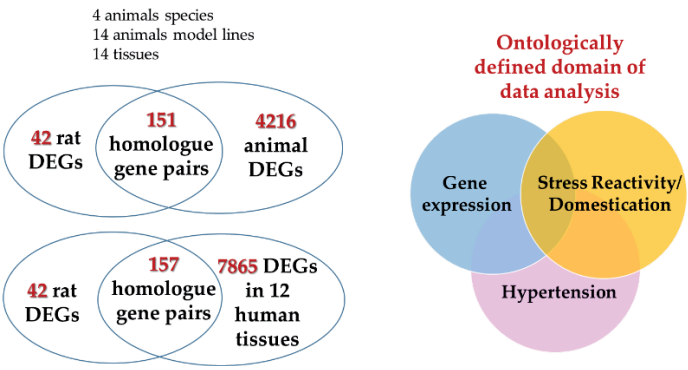


Fig. 1. Flowchart for ontological analysis of big transcriptomic data and differentially expressed genes in association with stress reactivity and hypertension

Analysis of these dependencies by the principal component method made it possible to obtain expressions for two components, one of which corresponded to arterial hypertension, and the second to stress reactivity. At the last stage, for each of the identified 42 DEGs in the hippocampus of tame and aggressive rats and the corresponding homologous human DEGs, the number of homologous pairs of DEGs with the same signs of the principal components, calculated by the $\log_2$ value of expression changes and their statistical significance, was determined. Identification of genes correlated with PC1 has identified two molecular markers associated with the risk of developing stress-sensitive hypertension: the β-protocadherin gene *Pcdhb9* (overexpression of which narrows the internal diameter of blood vessels) and the hemoglobin β-subunit gene, *Hbb-b1* (overexpression of which increases viscosity blood) [2].

We then focused on PAG to investigate molecular markers of age-related diseases (ARD). A large sample of different domestic animals shows an accumulation of genetic mutations that contribute to susceptibility to such diseases. Similar ontological analysis and principal component analysis of a big dataset identified only one gene out of 39 DEGs as a molecular marker associated with the risk of developing ARD. In particular, the *Fcgr2b* gene in the PAG of tame and aggressive rats correlated with susceptibility to ARD (Table 1). Increased expression of this gene aggravates immunogenic diseases associated with cell aging [3].

Table 1. Comparison of the effects of unidirectional changes in the expression of the human *FCGR2B* gene on age-related diseases (ARD) severity in humans and of its animal homologs on the microevolutionary events leading to domestic and wild animals

Effect of changes in human FCGR2B expression on ARD		Effect of changes in the expression of animal genes homologous to human FCGR2B					
Decreased expression	Increased expression	Decreased expression	Increased expression	Tissue	DEG	$\log_2$	P_{adj}
Aggravating age-related diseases: In South Asia and Africa, the human *FCGR2B*-deficient alleles have the most frequent occurrence as protectors against infection, susceptibility to which increases with age as immuno-senescence	Alleviating age-related diseases: *FCGR2B* has been explored using the "C-type lectin-like molecule-1"/"Fc-domain" fusion protein as a target antigen for chemotherapy against acute myeloid leukemia as a cellular-senescence-related immunogenic disease	Aggressive rat	Tame rat	Periaqueductal gray	*Fcgr2b*	2.02	0.05
				Hypothalamus	*Fcrl2*	1.12	0.05
					Fcgr3a	2.06	10^{-2}
				Midbrain Tegmentum	*Fcgr2b*	2.01	0.05
		Aggressive fox	Tame fox	Pituitary	*Fcrl1*	0.43	10^{-2}
		Wild rabbit	Domestic rabbit	Parietal-temporal cortex	*Fcgr3b*	1.35	10^{-2}

The transcriptomes of domestic and wild animals were processed accordingly and compared with the transcriptome of the hypothalamus of tame and aggressive gray rats. Here we also identified a number of genes correlated with principal components that reflect the effects of artificial selection and intraspecific variation [1].

Finally, we analyzed the 4th transcriptome, MTg, in the light of the influence of transcription factors on gene expression. Using bioinformatics methods, the correspondence between the increased level of expression of the *Ascl3* gene in tame rats and changes in the expression of genes from the list of DEGs in the MTg was analyzed. It was shown that motifs corresponding to potential binding sites for the transcription factor ASCL3 constitute the majority of all significantly enriched motifs for genes with reduced expression in tame animals. These results led us to the conclusion that the period of neurogenesis in tame rats is more extended than in aggressive ones.

Conclusion: Using an ontological approach and principal component analysis, we identified molecular markers of hypertension and age-related diseases, and also identified genes associated with domestication. In addition, it was shown that motifs corresponding to potential binding sites for the transcription factor ASCL3 constitute a significant part of all significantly enriched motifs for genes with reduced expression in tame animals, which indicates prolonged neurogenesis.

Funding: The study is supported by Russian government projects FWNR-2022-0015 and Kurchatov Genomic Center of ICG SB RAS agreement number 075-15-2019-1662. For data analysis, computational resources of the Bioinformatics CPC were used with the support of budget project No. FWNR-2022-0020.

References

1. Chadaeva et al. Domestication explains two-thirds of differential gene expression variance between domestic and wild animals; the remaining one-third reflects intraspecific and interspecific variation. *Animals*. 2021;11:2667. doi 10.3390/ani11092667
2. Oshchepkov et al. Stress reactivity, susceptibility to hypertension, and differential expression of genes in hypertensive compared to normotensive patients. *Int J Mol Sci.* 2022;23(5):2835. doi 10.3390/ijms23052835
3. Shikhevich et al. Differentially expressed genes and molecular susceptibility to human age-related diseases. *Int J Mol Sci*. 2023;24(4):3996. doi 10.3390/ijms24043996
4. Oshchepkov et al. Transcription factors as important regulators of changes in behavior through domestication of gray rats: quantitative data from RNA sequencing. *Int J Mol Sci.* 2022;23(20):12269. doi 10.3390/ijms232012269

doi 10.18699/bgrs2024-1.4-25

Differentially expressed genes and molecular susceptibility to human age-related diseases

Oshchepkov D.[1, 4]*, Shikhevich S.[1], Chadaeva I.[1], Khandaev B.[1, 2], Kozhemyakina R.[1], Zolotareva K.[1, 2], Kazachek A.[1, 2], Bogomolov A.[1, 2], Klimova N.V.[1], Ivanisenko V.A.[1, 2], Demenkov P.[1], Mustafin Z.[1, 2], Markel A.[1, 2], Savinkova L.[1], Kolchanov N.A.[1, 2], Kozlov V.[3], Ponomarenko M.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Research Institute of Fundamental and Clinical Immunology (RIFCI), SB RAS, Novosibirsk, Russia*

[4] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *diman@bionet.nsc.ru*

Key words: RNA-seq; *Rattus norvegicus*; age-related disease; bootstrap; correlation; differentially expressed gene; human; meta-analysis; molecular marker; principal component; qPCR

Motivation and Aim: Mainstream transcriptome profiling of susceptibility versus resistance to age-related diseases (ARDs) is focused on differentially expressed genes (DEGs) specific to gender, age, and pathogeneses. This approach fits in well with predictive, preventive, personalized, participatory medicine and helps understand how, why, when, and what ARDs one can develop depending on their genetic background. Within this mainstream paradigm, we wanted to find out whether the known ARD-linked DEGs available in PubMed can reveal a molecular marker that will serve the purpose in anyone's any tissue at any time. Recently we have obtained and compared transcriptomes of the midbrain tegmentum [1] in gray rats of a tame and an aggressive outbred strain; this allowed us to identify, in *in silico* settings, potential molecular markers for neoteny, which has a power to reverse ARDs. Furthermore, we have recently profiled transcriptomes in the hippocampus [2] of tame versus aggressive rat strains, in which deficient beta-protocadherins and beta-hemoglobin were found to be the statistically significantly most common theranostic molecular markers for ARD-related hypertension. That is why we sequenced, in the same way, one more transcriptome, that of the periaqueductal gray matter (PAG), in the tame versus aggressive rats, identified the corresponding behavior-related PAG-associated DEGs in them. We compared these DEGs with their homologous PubMed-based ARD-linked DEGs in animals and humans to find out whether there are invariant molecular markers for such diseases among them. *Methods and Algorithms*: We retrieved all PubMed-based ARD-linked DEGs in animals and humans, 37834 and 14535 DEGs respectively. We sequenced PAG transcriptome of tame versus aggressive rats, identified rat-behavior-related DEGs, and compared them with their known homologous animal ARD-linked DEGs. This analysis yielded statistically significant correlations between behavior-related and ARD-susceptibility-related fold changes (log2 values) in the expression of these DEG homologs. We found principal components, PC1 and PC2, corresponding to the half-sum and the half-difference of these log2 values, respectively. With the DEGs linked to ARD susceptibility and ARD resistance in humans used as controls, we verified these principal components. Finally, for each of the identified 39 DEGs in the PAG of tame and

aggressive rats and the corresponding homologous DEGs of humans and animals, we determined the number of homologous pairs of DEGs that have the same signs for the main components, calculated using the corresponding log2 values of changes expression, and their statistical significance.

Results: This analysis yielded only one statistically significant common molecular marker for ARDs: an excess of Fc receptor IIb suppressing immune cell hyperactivation. That is consistent with independent RNA-Seq data on the rise in the levels of murine Fcgr2b in astrocytes with age [4].

Conclusion: Finally, we propose the human immunoregulatory genes FCGR1A, FCGR2A, FCGR2B, FCGR2C, FCGR3A, and FCGR3B, homologous to the rat Fcgr2b gene and expressed in the majority of the human tissues, as theranostic molecular markers of age-related diseases.

Funding: The study is supported by the Government Budget Project FWNR-2022-0020 and computational resources of HPC facilities at collaborative center “Bioinformatics” ICG SB RAS.

References

1. Oshchepkov D., Chadaeva I., Kozhemyakina R., Shikhevich S., Sharypova E., Savinkova L., Klimova N.V., Tsukanov A., Levitsky V.G., Markel A.L. Transcription factors as important regulators of changes in behavior through domestication of gray rats: quantitative data from rna sequencing. *Int J Mol Sci.* 2022;23(20):12269. doi 10.3390/ijms232012269
2. Oshchepkov D., Chadaeva I., Kozhemyakina R., Zolotareva K., Khandaev B., Sharypova E., Ponomarenko P., Bogomolov A., Klimova N.V., Shikhevich S., Redina O., Kolosova N.G., Nazarenko M., Kolchanov N.A., Markel A., Ponomarenko M. Stress reactivity, susceptibility to hypertension, and differential expression of genes in hypertensive compared to normotensive patients. *Int J Mol Sci.* 2022;23(5):2835. doi 10.3390/ijms23052835
3. Shikhevich S., Chadaeva I., Khandaev B., Kozhemyakina R., Zolotareva K., Kazachek A., Oshchepkov D., Bogomolov A., Klimova N.V., Ivanisenko V.A., Demenkov P., Mustafin Z., Markel A., Savinkova L., Kolchanov N.A., Kozlov V., Ponomarenko M. Differentially expressed genes and molecular susceptibility to human age-related diseases. *Int J Mol Sci.* 2023;24(4):3996. doi 10.3390/ijms24043996
4. Clarke L.E., Liddelow S.A., Chakraborty C., Münch A.E., Heiman M., Barres B.A. Normal aging induces A1-like astrocyte reactivity. *Proc Natl Acad Sci USA.* 2018;115(8):E1896-E1905. doi 10.1073/pnas.1800165115

doi 10.18699/bgrs2024-1.4-26

Liver transcriptomics revealed species-specific responses during infections with three foodborne trematodes

Pakharukova M.Y.[1, 2]*, Lishai E.A.[1, 2], Zaparina O.G.[1], Kapushchak Y.K.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *pmaria@yandex.ru*

Key words: liver transcriptome; hamster model; helminths

Motivation and Aim: Trematodes of the Opisthorchiidae family vary in their carcinogenicity to humans. *Opisthorchis viverrini* and *Clonorchis sinensis* are classified as group 1A biological carcinogens, while *O. felineus* is classified as group 3A, a non-carcinogen [1, 2]. Comparison of three closely related species with different carcinogenic potential seems to be a promising model for studying the key mechanisms in liver metabolic shifts during biological carcinogenesis and identifying the cascade of regulatory signaling events and the dynamics of pre-carcinogenic changes.

The aim of the study was to identify differences in the activation of genes and signaling pathways in the liver of *Mesocricetus auratus* hamsters infected with three trematode species.

Methods and Algorithms: The following research methods were applied (RNA isolation, cDNA library construction, Western blot analysis, histological analysis of liver sections). Liver transcriptomes of golden hamsters (HiSeq Illumina, 2X150 bp) were sequenced at 1 and 3 months postinfection. Samples for the study were prepared in an SPF Animal Facility (free from specific pathogens) and analyzed under the same conditions. This made it possible to minimize comparison errors, such as a human factor, climate differences, specific pathogen infections of laboratory animals, etc.

Data processing was carried out using the following bioinformatics approaches: analysis of differential gene expression, functional enrichment of differentially expressed genes (DEGs), analysis of cellular composition using sequencing data of single liver cells, analysis of weighted gene coexpression networks by means of FastQC, STAR and R packages (DESeq2, cluster-Profiler, bisqueRNA, PCAtools, WGCNA). DEG enrichment analysis was performed using Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Molecular Signatures Database (MsigDB) databases.

Results: Sequencing of 24 DNA libraries resulted in an average of 37 million sequences per library. *C. sinensis* infection caused changes in the expression of the largest number of DEG compared to the uninfected animals – 3148 genes; *O. viverrini* infection – 1464 genes were differentially expressed, and *O. felineus* infection – 1408 genes.

All infections are characterized by enrichment in the inflammatory response signaling pathway, fibrogenesis and cell proliferation, IL2-STAT5, TNF-α NFkB, TGF-β, Hippo, MAPK, PI3K-Akt signaling pathways. However, species-specific response to each infection were also found. Thus, *O. viverrini* infection was characterized by significant enrichment in the FoxO and hypoxia pathways, *C. sinensis* infection – P53 and relaxin signaling pathway, *O. felineus* infection – TNF-α and ErbB signaling pathways.

The contribution of individual liver cells to transcriptomic response also depended on the type of infection. *C. sinensis* caused the most pronounced changes in the response of

various types of liver cells, including pro-inflammatory macrophages and stellate cells, which was confirmed using Western blot analysis for the content of cell marker proteins alpha smooth muscle actin (aSMA) and metalloproteinase 9 (MMP9).

Semi-quantitative histological analysis of hamster liver revealed hepatobiliary lesions and severity of damage during infections, which differed significantly between groups of infected animals. Infections with *C. sinensis* and *O. felineus* are most characterized by the profibrotic changes. Whereas *O. viverrini* infection is characterized by more pronounced neoplasia of the bile duct epithelium, which is considered a precancerous state. In general, the transcriptome data are consistent with the results of semi-quantitative histological analysis, where during infection with *C. sinensis*, liver periductal fibrosis was most pronounced, in the formation of which stellate cells are mostly involved.

We also analyzed the correlation (Pearson coefficient) between clusters of co-expressed genes and structural changes in the liver, such as fibrosis and neoplastic changes in the biliary epithelium using analysis of weighted gene co-expression networks.

Conclusion: This study was the first to conduct a comparative analysis of gene expression profiles in the liver of experimental animals infected with trematodes with different carcinogenic potential *O. viverrini*, *O. felineus* and *C. sinensis*. The studied trematodes have a species-specific effect on the hepatobiliary system, influencing changes in the expression of various genes and activation or suppression of cellular signaling pathways, which leads to differences in the severity of structural lesions to the liver and can explain the carcinogenic potential of closely related foodborne trematodes.

Funding: The study is supported by the Russian Science Foundation (No. 24-44-00048).

References

1. IARC. A review of human carcinogens. IARC working group on the evaluation of carcinogenic risks to humans. Monographs evaluation carcinogenic risks to human. *Biol Agents*. 2012;100B:1-441
2. Pakharukova M.Y., Mordinov V.A. The liver fluke Opisthorchis felineus: biology, epidemiology and carcinogenic potential. *Trans R Soc Trop Med Hyg*. 2016;110(1):28-36. doi 10.1093/trstmh/trv085

doi 10.18699/bgrs2024-1.4-27

Transcriptional activity of chicken W-linked genes at the lampbrush chromosome stage based on RNA-seq and RNA-FISH data

Plotnikov V.[1], Kulikova T.[1], Fedorov A.[1], Schelkunov M.[2, 3], Fedotova A.[2, 4], Krasikova A.[1]*

[1] *Saint-Petersburg State University, St. Petersburg, Russia*

[2] *Genomics Core Facility, Skolkovo Institute of Science and Technology, Moscow, Russia*

[3] *Institute for Information Transmission Problems, Moscow, Russia*

[4] *Lomonosov Moscow State University, Moscow, Russia*

* *alla.krasikova@gmail.com*

Key words: lampbrush chromosomes; maternal RNA; oocyte transcriptome; RNA-FISH; sex chromosomes

Motivation and Aim: In the domestic chicken, the W chromosome is annotated with 28 protein-coding genes, all of which have corresponding homologs on the Z chromosome. The expression of these genes has been documented in various tissues [1]. Notably, female germinative stem cells exhibit an abnormal increase in the expression of 21 Z-W gene pairs, implying the operation of distinct sex determination mechanisms [2]. The aim of this study is to investigate whether the expression of W chromosome genes persists in growing oocytes during the diplotene stage of prophase I of meiosis.
Methods and Algorithms: Transcription of W chromosome genes was examined by analysing RNA sequencing data, aligned to the *GGswu1* genome assembly, from both nuclear and cytoplasmic total and poly(A) RNA fractions extracted from chicken lampbrush stage oocytes [3]. The nuclear and cytoplasmic RNA-seq profiles were visually inspected in the Integrative Genome Viewer to assess the transcriptional activity of Z-W gene pairs. Transcriptional activity was estimated by read coverage along exons in the cytoplasmic RNA fraction, and also by read coverage along exons and introns in the nuclear RNA fraction. To verify the transcription of W chromosome genes, we performed RNA fluorescence *in situ* hybridisation (RNA-FISH) on isolated lampbrush chromosome preparations. Probes for RNA-FISH were prepared by PCR with primers for specific regions of the selected W chromosome genes.
Results: We demonstrated that the most Z-W gene pairs are expressed in chicken oocytes at the lampbrush chromosome stage according to RNA-seq data of oocyte cytoplasmic total and poly(A) RNA. Nuclear RNA-seq profile along exons and introns allows differentiation of transcripts originating from W-linked genes from their homologues on chromosome Z. For example, the *hnRNPK* gene is transcribed from both the Z and W chromosome loci, the full-length *RASA1* transcript is synthesised from the Z chromosome, *HINT1* gene is transcribed from Z chromosome, whereas the multicopy gene *HINT1W* (as a part of tandemly repeated sequence sate-HINT1) is not transcribed, and finally the *TCF4* gene appears to be transcriptionally silent on both sex chromosomes. The transcription of several genes on chicken lampbrush chromosome W was confirmed by RNA-FISH. A novel approach has been developed to synthesise a set of PCR-derived probes, eliminating the need for BAC clone-based probes. As a result, we clarified the RNA sequencing data for a number of W chromosome genes, such as

hnRNPK. Finally, we analysed the transcriptome of embryos at an early developmental stage (EGK.III) and revealed the inheritance of spliced mRNA from the Z-W gene pairs. *Conclusion*: In the current study, we found that the transcriptional activity of W chromosome genes observed in germinative stem cells is maintained in chicken oocytes at the lampbrush chromosome stage. The subsequent detection of spliced mRNA for Z-W gene pairs in early embryogenesis suggests their potential role in development as maternal factors. Remarkably, despite the extensive expansion of repetitive sequences in over 85 % of the W chromosome and enrichment of heterochromatin markers such as DNA hypermethylation, a number of W-linked genes are transcribed at the lampbrush chromosome stage of oogenesis.

Funding: The research was carried out using the equipment of the Genomics Core Facility (Skoltech) and the Molecular and Cell Technologies Resource Center (St. Petersburg State University).

References

1. Bellott D.W., Skaletsky H., Cho T.-J., Brown L., Locke D., Chen N. et al. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nat Genet*. 2017;49:387-394. doi 10.1038/ng.3778
2. Ichikawa K., Nakamura Y., Bono H., Ezaki R., Matsuzaki M., Horiuchi H. Prediction of sex-determination mechanisms in avian primordial germ cells using RNA-seq analysis. *Sci Rep*. 2022;12:13528. doi 10.1038/s41598-022-17726-7
3. Krasikova A., Kulikova T., Schelkunov M., Makarova N., Fedotova A., Berngardt V., Maslova A., Fedorov A. The first chicken oocyte nucleus whole transcriptomic profile defines the spectrum of maternal mRNA and non-coding RNA genes transcribed by the lampbrush chromosomes. *bioRxiv*. 2024. doi 10.1101/2024.02.05.577752

doi 10.18699/bgrs2024-1.4-28

Promoters of genes encoding β-amylase, albumin, and globulin in food plants have weaker affinity for TATA-binding protein as compared to non-food plants*: in silico* analysis

Podkolodnaya O.[1]*, Chadaeva I.[1], Zolotareva K.[1], Podkolodnyy N.[1, 3], Khandaev B.[1, 2], Rasskazov D.[1], Bogomolov A.[1, 2], Kazachek A.[1, 2], Vishnevsky O.[1, 2], Sharypova E.[1], Ponomarenko P.[1], Savinkova L.[1], Zemlyanskaya E.[1, 2], Ponomarenko M.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

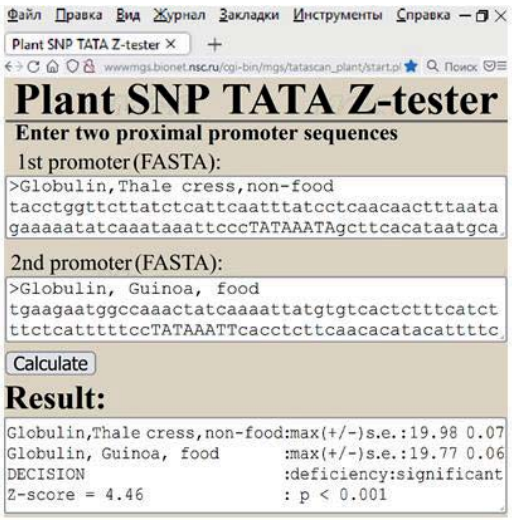
* *opodkol@bionet.nsc.ru*

Key words: food, allergen; albumin; globulin; β-amylase; gene; promoter; TATA-binding protein; TATA box; common wheat; *Triticum aestivum* L. (1753); plants; spontaneous artificial selection

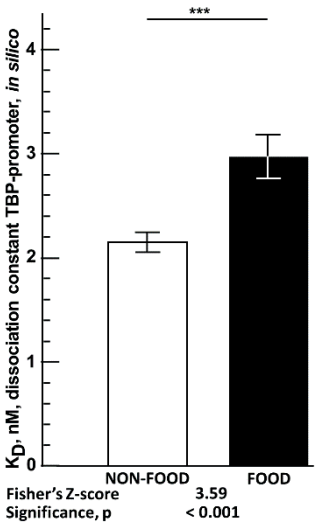
Motivation and Aim: It is generally accepted that during the domestication of food plants, the selection was focused on their productivity, the ease of their technological processing into food, and resistance to pathogens and environmental stressors. Besides, the palatability of plant foods and their health benefits could also be subjected to selection by humans in the past. Nonetheless, it is unclear whether in antiquity, aside from positive selection for beneficial properties of plants, humans also simultaneously selected against such detrimental properties as allergenicity. This topic is becoming increasingly relevant as the allergization of population grows, being a major challenge for modern medicine. That is why, breeders are constantly trying to create hypoallergenic forms of food plants. *Methods and Algorithms*: Accordingly, in this paper, albumin, globulin, and β-amylase of common wheat *Triticum aestivum* L. (1753) are analyzed, which have been identified earlier as targets for attacks by human class E immunoglobulins [1]. At the genomic level, we wanted to find signs of past negative selection against the allergenicity of these three proteins (albumin, globulin, and β-amylase) during the domestication of ancestral forms of modern food plants. Using our public web service Plant_SNP_TATA_Z-tester [2], here we focused the search on the TATA-binding protein (TBP)-binding site because it is located within a narrow region (between positions –70 and –20 relative to corresponding transcription start sites), is the most conserved, necessary for primary transcription initiation, and is the best-studied regulatory genomic signal in eukaryotes. By this way, we have estimated the equilibrium dissociation constant (K_D) of TBP complexes with plant proximal promoters (as its output data) using 90 bp of their DNA sequences (as its input data) for 235 gene promoters representing 28 plant species, as described in Table 1 and illustrated in Fig. 1a.

Table 1. Characteristics of 235 nucleotide sequences of proximal promoters of food plant genes homologous to the studied globulin (*Glo*), albumin (*Alb*), and β-amylase (*Bmy*) gens from common wheat *Triticum aestivum* L. (1753)

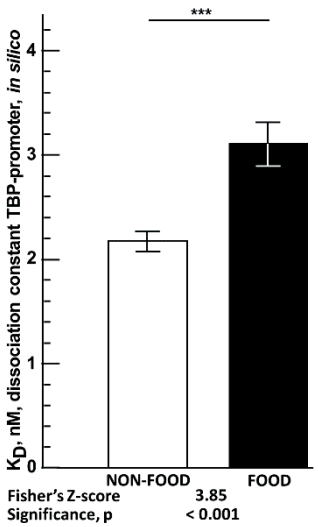
Food plant species		Number of promoters		
#	Name	*Glo*	*Alb*	*Bmy*
1	buckwheat *Fagopyrum esculentum* Moench, 1794	1		
2	maidenhair tree *Ginkgo biloba* L., 1771	1		
3	Yoshino cherry *Prunus yedoensis* var. nudiflora Koehne, 1912	1		2
4	maize (*Zea mays L., 1753*	1		
5	oat *Avena sativa* L., 1753	2		
6	waxberry *Morella rubra* Siebold & Zucc.	2	2	1
7	quinoa *Chenopodium quinoa* Willd., 1798	2		
8	rice *Oryza sativa* L., 1753	3		
9	melon *Cucumis melo* L., 1753	4	2	6
10	cardoon *Cynara cardunculus* L.	4		
11	cork oak *Quercus suber* L.	4		
12	wine grape *Vitis vinifera* L.	9		9
13	Congolese coffee *Coffea canephora* Pierre ex A.Froehner, 1897	1		
14	pepper *Capsicum annuum* L., 1753	26	8	
15	sesame *Sesamum indicum* L.		1	
16	kiwifruit nashi-kazura *Actinidia rufa* Franch. & Sav		1	27
17	Brazil nut *Bertholletia excelsa* Humb. & Bonpl.		1	
18	soybean *Glycine max* (L.) Merr., 1917		2	
19	pea *Pisum sativum* L., 1753		4	
20	perilla *Perilla frutescens* var. *hirtella* (Nakai) Makino		5	
21	almond *Prunus dulcis* (Mill.) D.A.Webb, 1967		8	4
22	mandarin unshiu *Citrus unshiu* (Tanaka ex Swingle) Marcow., 1921		15	
23	tea *Camellia sinensis* (L.) Kuntze, 1887			1
24	barley *Hordeum vulgare* L. (1753)			2
25	hibiscus *Hibiscus syriacus* L. (1753)			2
26	pineapple *Ananas comosus* (L.) Merr., 1917			3
27	olive *Olea europaea* L., 1753			4
28	sweet wormwood *Artemisia annua* L.	13	35	16
TOTAL number of food plant species		**15**	**12**	**12**



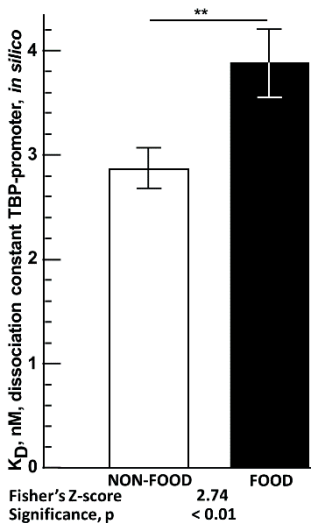
(a)



(b)



(c)



(d)

Fig. 1. The statistically significant difference between the studied food plants and non-food plants in the *in silico* estimates of K_D for complexes of plant TBP with 90 bp proximal promoters of their genes encoding globulins (b), albumins (c) and β-amylases (d), *as calculated by our publicly available Web service Plant_SNP_TATA_Z-tester [2] (a). Legend*: Bar height, the arithmetic mean of the equilibrium dissociation constant (K_D) for complexes between plant TBP and the 90 bp proximal promoters of the indicated plant genes; error bars, standard error of the mean (SEN); asterisks, statistical significance according to Fisher's Z-test, namely: ** for $p < 0.01$ and *** for $p < 0.001$

Results: It was found that compared with non-food plants, food plants are characterized by significantly weaker affinity of TBP for proximal promoters of their genes homologous to the genes of common-wheat globulin, albumin, and β-amylase (food allergens), as presented in Fig. 1b, 1c and 1d: $p < 0.01$ according to Fisher's Z-test [3].
Conclusion: This suggests that in the past, humans to reduce the expression of food plant genes encoding these allergenic proteins carried out spontaneous artificial selection.
Funding: The study is supported by the Government Budget Project FWNR-2022-0020.

References

1. Wang Y., Weng J., Zhu C., Ai R., Zhou J., Wang C., Chen Q., Fu L. Allergenicity assessment and allergen profile analysis of different Chinese wheat cultivars. *World Allergy Organ J*. 2021;14(7):100559. doi 10.1016/j.waojou.2021.100559
2. Rasskazov D., Chadaeva I., Sharypova E. et al. Plant_SNP_TATA_Z-Tester: A Web service that unequivocally estimates the impact of proximal promoter mutations on plant gene expression. *Int J Mol Sci*. 2022;23(15):8684. doi 10.3390/ijms23158684
3. Vishnevsky O.V., Chadaeva I.V., Sharypova E.B. et al. Promoters of genes encoding β-amylase, albumin and globulin in food plants have weaker affinity for TATA-binding protein as compared to non-food plants: in silico analysis. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):798-805. doi 10.18699/VJGB-22-96

oi 10.18699/bgrs2024-1.4-29

Integration of genomic and clinical data: assessment of the influence of SNPs in the TATA-box region of human genes on changes in the expression of these genes and their associations with clinical and phenotypic observations

Podkolodnyy N.L.[1, 2]*, Filonov S.V.[1], Podkolodnaya O.A.[1], Tverdokhleb N.N.[1], Ponomarenko P.M.[1], Rasskazov D.A.[1], Bogomolov A.G.[1], Ponomarenko M.P.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

* *pnl@bionet.nsc.ru*

Key words: TATA box; affinity; TBP; single nucleotide polymorphism; database; genome-wide analysis; Human_SNP_TATAdb database; ClinVar database; candidate SNP markers of diseases

Motivation and Aim: Genome-wide association studies (GWAS) have shown that most SNPs that are significantly associated with disease susceptibility are located in non-coding regions [1–3], and more than 90 % of them are located in regulatory elements [4]. Now, one of the most studied regulatory regions is the TATA box region in the promoter, the sequence of which determines the affinity of the TBP protein (TATA binding protein), which is a key transcription initiation factor. Mutations in this region can affect the binding of the TBP protein to the promoter and, consequently, gene expression [5, 6]. Previously, the Institute of Cytology and Genetics SB RAS (ICG) has developed a method for predicting the affinity of TBP to gene promoters based on a three-step binding mechanism, including sliding of TBP along DNA, stopping of TBP at the binding site, and fixing the TBP-promoter complex due to bending of the DNA helix [7]. The method showed a high correlation of theoretical predictions with measured values during repeated experimental testing by independent groups of researchers. Based on this model, the ICG has developed the SNP_TATA_Z-tester and SNP_TATA_Comparator web services, which allow calculating a statistical estimation of the SNP-induced change in the affinity of TBP binding to the human gene promoter and predicting changes in expression that may be associated with a genetic predisposition to diseases or phenotypic features of the organism. Using this web service, we previously identified candidate SNP markers for autoimmune diseases [8], behavioral disorders [9], chronopathologies [10] and other diseases. In this work, we integrated into a database information about SNPs in human gene promoters, obtained by automatic extraction from various heterogeneous data sources, as well as the results of assessing the affinity of TBP to the promoter and the specificity of the TBP binding site using a three-step binding model and assessing their effect on gene expression for the reference genome promoters and promoters with SNPs as well as clinical and phenotypic observations associated with SNPs presented in the ClinVar database. This makes it possible to identify candidate genetic markers of various clinical and phenotypic manifestations among SNPs located in the TATA box region.

Methods and Algorithms: Data on genes and their attributes, transcription starts and transcripts were obtained from the Ensembl web service. To access the services and

databases used in the work, the Bioconductor library of the R language was used, with the following packages:
1. biomaRt is a package that provides an interface to the ENSEMBL collection of databases, allowing large volumes of data to be retrieved in a unified way and used in data analysis in Bioconductor.
2. BSgenome.Hsapiens.NCBI.GRCh38 is package that provides access to the Homo sapiens (Human) genome sequence provided by NCBI (GRCh38.p13).
3. SNPlocs.Hsapiens.dbSNP155.GRCh38 is a dbSNP 155 access package including information on 949,021,448 SNPs in chromosomes 1-22, X, Y and MT.
4. Information on the associations of snps with phenotypic and clinical observations presented in the ClinVar database (Apr, 22, 2024).
For each promoter, SNPs located within [–90,–1] from the start of transcription were identified. The affinity of TBP for DNA was calculated using a three-step binding model previously developed at the Institute of Cytology and Genetics SB RAS [7] and a multi-threaded high-performance version of the SNP_TATA_Z-tester program also implemented by us. This program also allows you to evaluate the statistical significance of changes in the affinity of the TBP protein for the promoter due to point nucleotide substitutions (SNPs) in the promoter using a z-test.
Results: A software package was developed to extract data from available information resources about all human genes and their transcripts, promoter sequences and SNPs localization in them. A multi-threaded version of the SNP_TATA_Comparator program was developed to assess changes in the affinity of TBP for DNA and a genome-wide analysis of the influence of known SNPs in the promoters of all human genes on changes in the expression of these genes was carried out. Integration of genomic data with clinical and phenotypic observations presented in the clinVar database was carried out. As a result, the database Human_SNP_TATAdb contains the following information [11]:
● 62,603 genes, of which 19,314 encode proteins.
● 117,414 transcripts, of which 63,141 encode proteins.
● 5,305,816 SNP variants in gene promoters in the [–90,–1] interval from the start of transcription, of which 3,199,285 are in the promoters of protein-coding genes.
● For 445,875 SNP variants in the promoter of a protein-coding gene, we predicted that they statistically significantly (p-value < 0.05) change the level of TBP affinity for this promoter. Among them, for 3,847 SNP variants, there are clinical and phenotypic observations presented in the ClinVar database.
The results of genome-wide data analysis are presented, including the features of the distribution of genes by the number of transcripts, the distribution of SNPs affecting the affinity of TBP to DNA by positions within promoters, as well as patterns linking the affinity of TBP to the promoter, the specificity of the TBP binding site to the promoter and other characteristics of promoters. The results of genome-wide analysis showed that the affinity of TBP to the promoter and the specificity of its binding site are statistically related to other characteristics of promoters important for the functional classification of promoters and the study of the features of differential gene expression.
Conclusion: This work presents the integrated database Human_SNP_TATAdb, which includes information on single nucleotide polymorphisms in human gene promoters, the results of assessing the affinity of TBP to the promoter using a three-step binding model, and assessing their impact on gene expression for wild-type promoters and promoters with single nucleotide polymorphism as well as clinical and phenotypic observations associated with snps presented in the clinvar database. The results of genome-wide

analysis showed that the affinity of TBP for the promoter and the specificity of its binding site are statistically associated with other characteristics of promoters that are important for the functional classification of promoters and the study of differential gene expression patterns.

We have shown that the affinity of the TBP protein to the promoter, the specificity of the TBP binding site to the promoter, and assessments of changes in these characteristics with single nucleotide polymorphisms presented in the Human_SNP_TATAdb database may be important for identifying candidate markers of genetic susceptibility to diseases, identifying and functional interpretation of classes of promoters that are similar in the mechanism of regulation of the early stage of transcription initiation, etc.

Funding: The study is supported by the Government Budget Projects (No. FWNR-2022-0020 and No. 0251-2022-0005).

References

1. Hindorff L.A., Sethupathy P., Junkins H.A., Manolio T.A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009;106(23):9362-9367. doi 10.1073/pnas.0903103106
2. French J.D., Edwards S.L. The role of noncoding variants in heritable disease. *Trends Genet*. 2020(36):880-891. doi 10.1016/j.tig.2020.07.004
3. Chandra V., Bhattacharyya S., Schmiedel B.J. et al. Promoter interacting expression quantitative trait loci are enriched for functional genetic variants. *Nat Genet.* 2021;(53):110-119. doi 10.1038/s41588-020-00745-3
4. Maurano M.T., Humbert R., Rynes E., Thurman R.E., Haugen E. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337(6099):1190-1195. doi 10.1126/science.1222794
5. Ravarani C., Chalancon G., Breker M. et al. Affinity and competition for TBP are molecular determinants of gene expression noise. *Nat Commun.* 2016;7:10417
6. Ponomarenko M., Rasskazov D., Arkova O. et al. How to Use SNP_TATA_Comparator to Find a Significant Change in Gene Expression Caused by the Regulatory SNP of This Gene's Promoter via a Change in Affinity of the TATA-Binding Protein for This Promoter. *Biomed Res Int*. 2015;2015:359835
7. Ponomarenko P.M., Savinkova L.K., Drachkova I.A., Lysova M.V., Arshinova T.V., Ponomarenko M.P., Kolchanov N.A. A Step-by-step Model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism. *Dokl Biochem Biophys*. 2008;419:88-92. doi 10.1134/S1607672908020117
8. Ponomarenko M.P. et al. Candidate SNP markers of genderbiased autoimmune complications of monogenic diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Front Immunol*. 2016a;7:130. doi 10.3389/fimmu.2016.00130
9. Chadaeva I.V., Ponomarenko M.P., Rasskazov D.A. et al. Candidate SNP markers of aggressiveness-related complications and comorbidities of genetic diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BMC Genomics*. 2016;17(Suppl. 14):995. doi 10.1186/s12864-016-3353-3
10. Ponomarenko P., Rasskazov D., Suslov V., Sharypova E., Savinkova L., Podkolodnaya O., Podkolodny N.L., Tverdokhleb N.N., Chadaeva I., Ponomarenko M., Kolchanov N. Candidate SNP markers of chronopathologies are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *Biomed Res Int*. 2016b;2016:8642703. doi 10.1155/2016/8642703
11. Filonov S.V., Podkolodnyy N.L., Podkolodnaya O.A., Tverdokhleb N.N., Ponomarenko P.M., Rasskazov D.A., Bogomolov A.G., Ponomarenko M.P. Human_SNP_TATAdb: a database of SNPs that statistically significantly change the affinity of the TATA-binding protein to human gene promoters: genome-wide analysis and use cases. *Vavilov J Genet Breed*. 2023;27(7):728-736. doi 10.18699/VJGB-23-85

doi 10.18699/bgrs2024-1.4-30

Bioinformatics of transcription initiation in eukaryotes: TATA binding protein (TBP) and TBP binding sites in gene promoters

Ponomarenko M.[1]*, Chadaeva I.[1], Zolotareva K.[1], Podkolodnyy N.[1, 2], Sharypova E.[1], Ponomarenko P.[1], Rasskazov D.[1], Bogomolov A.[1, 3], Savinkova L.[1], Kolchanov N.[1, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *pon@bionet.nsc.ru*

Key words: eukaryote, gene, promoter, TATA box; TATA-binding protein; SNP; *in silico* estimate; *in vivo* test; handmade annotation; target-assisted breeding

Motivation and Aim: One of the largest worldwide scientific projects now is “1000 Genomes”, which has already sequenced the individual genomes of many hundreds of thousands peoples. Their consensus sequence of the most frequent nucleotides to each position of genomic DNA is the reference human genome, difference of which from individual human genomes is single nucleotide polymorphism (SNP). The dbSNP database contains over one hundred of million experimentally proven SNPs [1]. The clinical search for biomedical SNP markers is the identification of SNP, the frequencies of which alleles in patients are statistically significantly different in comparison with healthy people that is obligatory criterion without alternative because of there is a chance to take such SNP-markers into account during the treatment of people. It is slowly, expensively and labor intensively applying to each of the all nine billion potential SNPs in humans. Preliminary computer-based analysis of unannotated SNPs can make this search for clinical SNP-markers much faster, cheaper, and more targeted due to rejecting neutral SNPs, which make up the absolute majority in humans according to both neutral evolution theory and Haldane's dilemma. The best studied is the SNP of protein-coding regions of genes, which disrupts the structure and function of proteins in the same way in all cells, tissues, and organs. That is why, these SNPs are easier to find, but their pathogenic effects cannot be avoid without genome editing, which is intensively studied in the case of human disease models using laboratory animals, as well as for plants and microorganisms, but is strictly prohibited in the case of humans. At the same time, the least studied is the regulatory gene region SNPs, which keep proteins in norm and change protein concentrations, which vary in cells, tissues, and organs for many reasons. That is why, they are more difficult to find, but its pathological effects can be corrected with medication and lifestyle changes. That is why, regulatory SNP-markers seem to be most promising for the current practice within genome-based personalized medicine. Among all regulatory SNPs, the best studied are those for the TATA-binding protein (TBP) binding sites, the canonical form of which is the TATA box. Most often TBP-sites can be found nearly 30 bp just in front of any transcription start sites of promoters as the obligatory regulatory element responsible for the primary initiation of transcription starting from just these sites, which is no longer required by subsequent re-initiation of transcription, but without which the phenomenon of transcription as such is impossible [2–4].

Methods and Algorithms: Therefore, in this review, we briefly recall how a three-step model of TBP binding to the promoters of human genes was proposed [5], which was then confirmed experimentally by independent authors [6]. Besides, we will describe how we have next created the Web-service SNP_TATA_Comparator [5] was, and how to use it in practice for genome-wide analysis of SNP within the human gene promoters. *Results*: As well, we present the genome-wide knowledge base Human_SNP_DATAdb [8], which among the all 11,384,651 SNPs within 90 bp proximal promoters of the all 17,079 protein-coding genes pinpoints to 602,547 SNPs altering significantly the TBP-promoter affinity and, thereby, the expression of the human genes under these promoters. *Conclusion*: Finally, Human_SNP_DATAdb has 57,612 annotations on how some of the above-said 602,547 candidate SNP-markers of the human gene expression changes can manifest in susceptibility to human diseases according to the PubMed database [9].
Funding: The study is supported by the Government Budget Project FWNR-2022-0020.

References

1. Day I.N. dbSNP in the detail and copy number complexities. *Hum Mutat*. 2010;31(1):2-4. doi 10.1002/humu.21149
2. Martianov I., Viville S., Davidson I. RNA polymerase II transcription in murine cells lacking the TATA binding protein. *Science*. 2002;298(5595):1036-1039. doi 10.1126/science.1076327
3. Auble D.T. The dynamic personality of TATA-binding protein. *Trends Biochem Sci*. 2009;34(2):49-52. doi 10.1016/j.tibs.2008.10.008
4. Rhee H.S., Pugh B.F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*. 2012;483(7389):295-301. doi 10.1038/nature10799
5. Ponomarenko P.M., Savinkova L.K., Drachkova I.A., Lysova M.V., Arshinova T.V., Ponomarenko M.P., Kolchanov N.A. A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism. *Dokl Biochem Biophys*. 2008;419:88-92. doi 10.1134/s1607672908020117
6. Delgadillo R.F., Whittington J.E., Parkhurst L.K., Parkhurst L.J. The TATA-binding protein core domain in solution variably bends TATA sequences via a three-step binding mechanism. *Biochemistry*. 2009;48(8):1801-1809. doi 10.1021/bi8018724
7. Ponomarenko M., Rasskazov D., Arkova O., Ponomarenko P., Suslov V., Savinkova L., Kolchanov N. How to use SNP_TATA_Comparator to find a significant change in gene expression caused by the regulatory SNP of this gene's promoter via a change in affinity of the TATA-binding protein for this promoter. *Biomed Res Int*. 2015;2015:359835. doi 10.1155/2015/359835
8. Filonov S.V., Podkolodnyy N.L., Podkolodnaya O.A., Tverdokhleb N.N., Ponomarenko P.M., Rasskazov D.A., Bogomolov A.G., Ponomarenko M.P. Human_SNP_TATAdb: a database of SNPs that statistically significantly change the affinity of the TATA-binding protein to human gene promoters: genome-wide analysis and use cases. *Vavilov J Genet Breed*. 2023;27(7):728-736. doi 10.18699/VJGB-23-85
9. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)*. 2011;2011:baq036. doi 10.1093/database/baq036

doi 10.18699/bgrs2024-1.4-31

Candidate SNP markers significantly altering the affinity of TATA-binding protein for the promoters of human hub genes for atherogenesis, atherosclerosis and atheroprotection

Tverdokhleb N.[1]*, Chadaeva I.[1], Filonov S.[1], Zolotareva K.[1], Podkolodnyy N.[1, 3], Khandaev B.[1, 2], Rasskazov D.[1], Bogomolov A.[1, 2], Kazachek A.[1, 2], Oshchepkov D.[1, 2], Ivanisenko V.[1, 2], Demenkov P.[1], Podkolodnaya O.[1], Ponomarenko P.[1], Mustafin Z.[1, 2], Kondratyuk E.[1, 4, 5], Savinkova L.[1], Ponomarenko M.[1], Suslov V.[1], Kolchanov N.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

[4] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[5] *Siberian Federal Scientific Centre of Agro-BioTechnologies, RAS, Krasnoobsk, Novosibirsk region, Russia*

* *nata@bionet.nsc.ru*

Key words: atherogenesis; atheroprotection; atherosclerosis; hub gene; promoter; TBP; TATA box; SNP; candidate SNP marker; gene expression change; natural selection; *in silico* verification

Motivation and Aim: Atherosclerosis is a systemic disease in which focal lesions in arteries promote the build-up of lipoproteins and cholesterol they are transporting. The development of atheroma (atherogenesis) narrows blood vessels, reduces the blood supply and leads to cardiovascular diseases. According to the World Health Organization (WHO), cardiovascular diseases are the leading cause of death, which has been especially boosted since the COVID-19 pandemic. There is a variety of contributors to atherosclerosis, including lifestyle factors and genetic predisposition. Antioxidant diets and recreational exercises act as atheroprotectors and can retard atherogenesis. The search for molecular markers of atherogenesis and atheroprotection for predictive, preventive and personalized medicine appears to be the most promising direction for the study of atherosclerosis.

Methods and Algorithms: Here, we have analyzed 1068 human genes associated with atherogenesis, atherosclerosis and atheroprotection [1], as depicted in the Figure 1.

Results: The hub genes regulating these processes have been found to be the most ancient by means of the an *in silico* phylostratigraphic analysis [2, 3]. *In silico* analysis using our freely available web service SNP_TATA_Comparator [4] of all 5112 SNPs in their promoters has revealed 330 candidate SNP markers, which statistically significantly change the affinity of the TATA-binding protein (TBP) for these promoters, as shown in the Table 1.

Conclusion: These molecular markers have made us confident that natural selection acts against underexpression of the hub genes for atherogenesis, atherosclerosis and atheroprotection, while upregulation of the only atheroprotection-related genes promotes human health according to the Table 2.

Funding: The study is supported by the Government Budget Project FWNR-2022-0020.

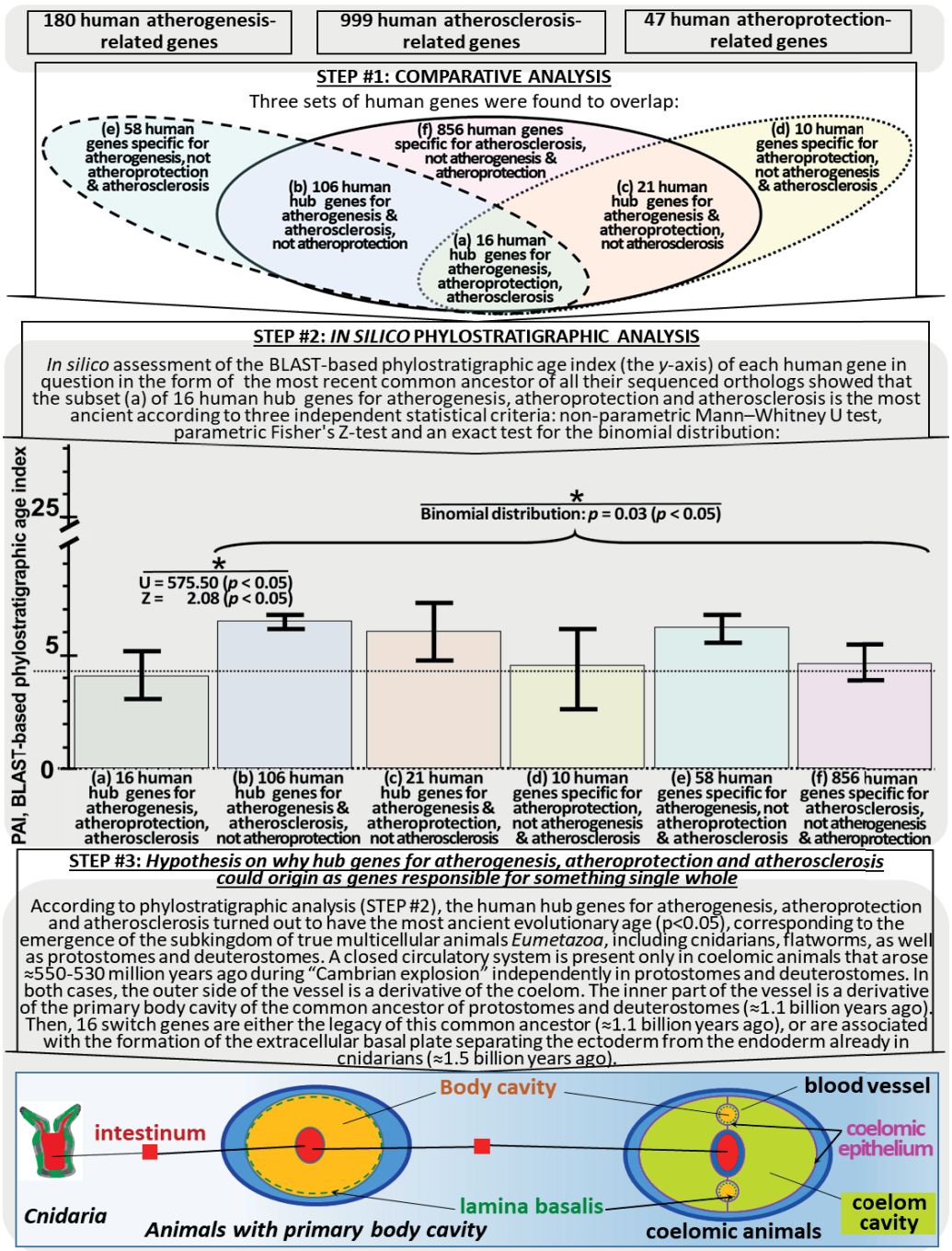


Fig. 1. The flowchart depicting step-by-step how we studied all 1068 human genes associated with atherogenesis, atherosclerosis and atheroprotection according to the NCBI Gene database [1]. *Legend*: PAI, genes' phylostratigraphic age index evaluated against the BLAST-based scale [2] using the freely available web service Orthoscape [3]. the heights of the gene bars and error bars, the arithmetic mean of the PAIs and the standard error of the mean (SEM); the dotted line, an auxiliary line running above the gray-green bar for the subset (a) of 16 human hub genes for atherogenesis, atherosclerosis and atheroprotection, the tops of the other five bars being above the line. (*): statistical significance $p < 0.05$. U and Z are the statistics in the nonparametric Mann–Whitney U test and parametric Fischer's Z-test according to Statistica (StatSoft^{TM}, USA)

Table 1. Candidate SNP makers significantly upregulating and downregulating TATA binding protein (TBP) affinity for the promoters of the human hub genes for atherogenesis, atherosclerosis and atheroprotection found in this work in comparison with those independently identified within whole human genome [5, 6] as an indicator of the neutral drift according to Haldane's dilemma [7] and neutral evolution theory [8]

Human Genome Assembly GRCh38/hg38, dbSNP Rel. 155	Number of objects ($N_{OBJECTS}$; "↑", upregulation; "↓", downregulationt)					Neutral drift: $p(H_0: N_{\downarrow} \geq 4N_{\uparrow})$ [7, 8]
	N_{GENE}	N_{SNP}	N_{MARKER}	$N_{\downarrow}$	$N_{\uparrow}$	Binomial distribution, *Bonferroni's correction*, P_{ADJ}
Whole-genome norm for SNPs significantly altering TBP-sites [5, 6]	30,000	10^5	1000	800	200	1.00
Human hub genes for athero-genesis, atheroprotection, and atherosclerosis [this work]	16	5112	466	150	316	10^{-4}

Table 2. Candidate SNP makers significantly changing TBP affinity for the promoters of the human hub genes for atherogenesis, atherosclerosis and atheroprotection

Atherosclerosis-related processes	The number of candidate SNP markers that have been found to significantly change TBP affinity for the promoters of the hub genes for atherogenesis, atherosclerosis and atheroprotection		Statistical significance binomial distribution, Bonferroni's correction, P_{ADJ}
	Human health failure	Human health improvement	
Atherogenesis	186	144	1.00
Atherosclerosis	195	135	0.19
Atheroprotection	**91**	**239**	**10^{-4}**

References

1. Brown G.R., Hem V., Katz K.S., Ovetsky M., Wallin C., Ermolaeva O., Tolstoy I., Tatusova T., Pruitt K.D., Maglott D.R., Murphy T.D. Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res*. 2015;43:D36-D42. doi 10.1093/nar/gku1055
2. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410. doi 10.1016/S0022-2836(05)80360-2
3. Mustafin Z.S., Lashin S.A., Matushkin Y.G., Gunbin K.V., Afonnikov D.A. Orthoscape: a cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles. *BMC Bioinformatics*. 2017;18(Suppl. 1):1427. doi 10.1186/s12859-016-1427-5
4. Ponomarenko M., Rasskazov D., Arkova O., Ponomarenko P., Suslov V., Savinkova L., Kolchanov N. How to use SNP_TATA_Comparator to find a significant change in gene expression caused by the regulatory SNP of this gene's promoter via a change in affinity of the TATA-binding protein for this promoter. *Biomed Res Int*. 2015;2015:359835. doi 10.1155/2015/359835
5. 1000 Genomes Project Consortium; Abecasis G.R., Auton A., Brooks L.D., DePristo M.A., Durbin R.M., Handsaker R.E., Kang H.M., Marth G.T., McVean G.A. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65. doi 10.1038/nature11632
6. Kasowski M., Grubert F., Heffelfinger C., Hariharan M., Asabere A. et al. Variation in transcription factor binding among humans. *Science*. 2010;328(5975):232-235. doi 10.1126/science.1183621
7. Haldane J.B.S. The cost of natural selection. *J Genet*. 1957;55:511-524
8. Kimura M. Evolutionary rate at the molecular level. *Nature*. 1968;217(5129):624-626. doi 10.1038/217624a0

doi 10.18699/bgrs2024-1.4-32

AtSNP_TATAdb: a knowledge base of uniform guess of how promoter proximal SNPs alter Arabidopsis gene expression

Zolotareva K.[1]*, Chadaeva I.[1], Khandaev B.[1, 2], Podkolodnyy N.[1, 3], Tverdokhleb N.[1], Ponomarenko P.[1], Sharypova E.[1], Filonov S.[1], Kondratyuk E.[1, 4, 5], Zemlyanskaya E.[1, 2], Rasskazov D.[1], Bogomolov A.[1, 2], Savinkova L.[1], Podkolodnaya O.[1], Kolchanov N.[1, 2], Ponomarenko M.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

[4] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[5] *Siberian Federal Scientific Centre of Agro-BioTechnologies, RAS, Krasnoobsk, Novosibirsk region, Russia*

** ka125699ri@yandex.ru*

Key words: plant; gene; promoter; TATA box; TATA-binding protein; SNP; expression alteration; genome-wide analysis; *in silico* estimate; *in vivo* test; handmade annotation; target-assisted breeding

Motivation and Aim: The mainstream of the post-genome target-assisted plant breeding is biofortification as high-throughput phenotyping along with genome-based selection.
Methods and Algorithms: Therefore, in this work, we used our previously developed Web-service Plant_SNP_TATA_Z-tester [1] to run a uniform *in silico* analysis of the transcription alterations of 54,013 protein-coding transcripts from 32,833 *Arabidopsis thaliana* L. genes caused by 871,707 single nucleotide polymorphisms (SNPs) within the 90 bp proximal promoter regions. We took their DNA sequences and SNPs from the databases Ensembl Plant [2] and TAIR [3], respectively, as depicted in Fig. 1.
Results: The analysis has identified 54,993 SNPs, each of which can significantly either upregulate or downregulate a proper Arabidopsis gene by means of an alteration in TATA-binding protein (TBP) binding affinity to the promoters carrying this SNP. The existence of these SNPs in highly conserved proximal promoters may be explained as intraspecific diversity kept by the stabilizing natural selection. To support this, using the database PubMed [4] we hand-annotated papers on some of the Arabidopsis genes possessing these SNPs or on their orthologs in other plant species and demonstrated the effects of changes in these gene expressions on plant vital traits. We integrated *in silico* estimates of the TBP-promoter affinity in the AtSNP_TATAdb knowledge base [5] (https://www.sysbio.ru/AtSNP_TATAdb/) and showed their significant correlations with independent *in vivo* experimental data, as exemplified in Fig. 2.
Conclusion: These correlations appeared to be robust to variations in statistical criteria, genomic environment of TATA box regions, plants species and growing conditions.
Funding: The study is supported by the Government Budget Project FWNR-2022-0020.

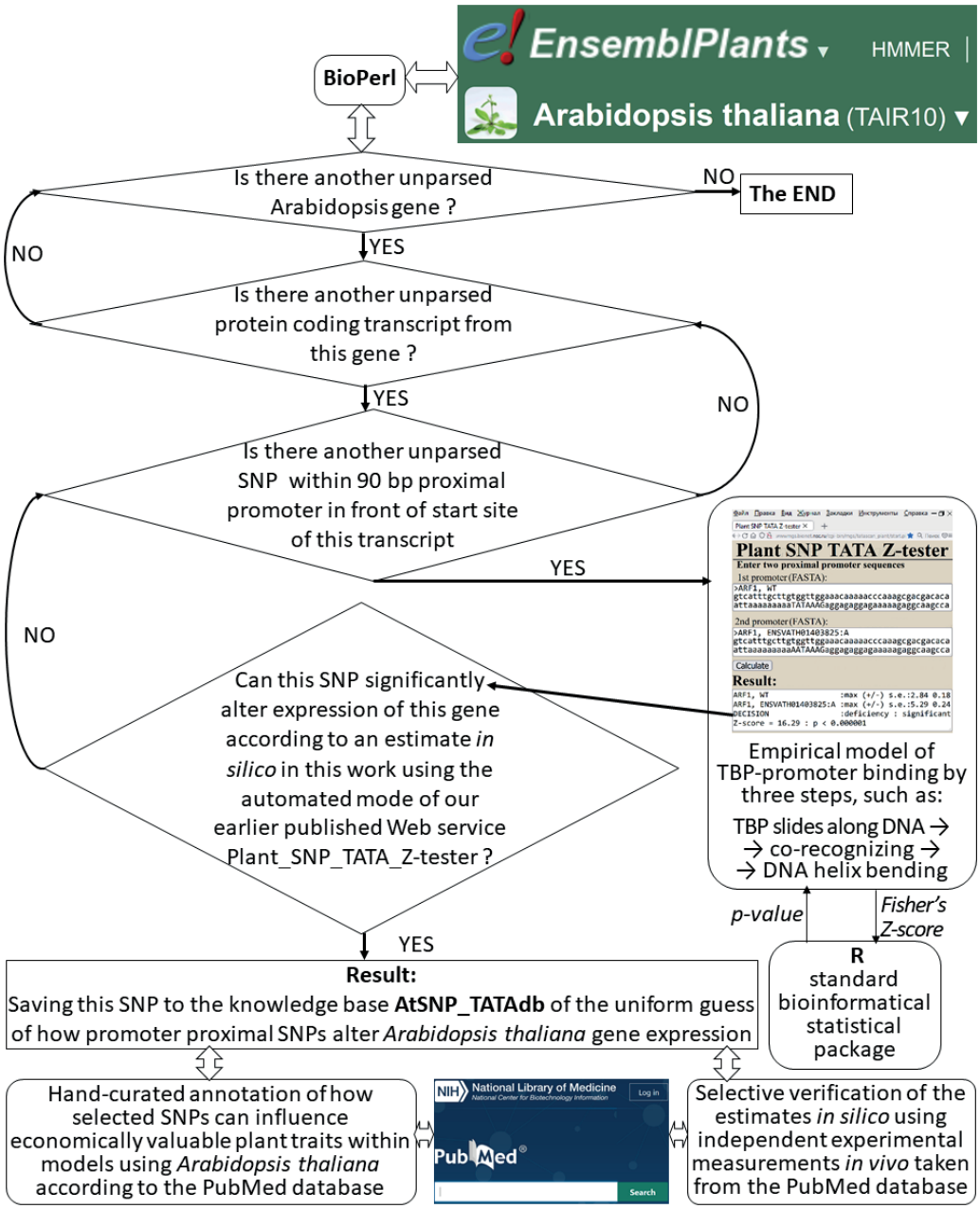


Fig. 1. Flowchart of AtSNP_TATAdb knowledge base [5] development by processing genome-wide information from both database Ensembl Plant [2] an TAIR [3] by first using Plant_SNP_TATA_Z-tester [1] and, after that, selectively annotating some of the SNPs by a search in the PubMed database [4] using information about where, when and under what conditions changes in expression of these genes or their homolog genes were observed in agricultural studies. Finally, we selectively verified estimates *in silico* using independent experimental data *in vivo* taken from the PubMed database [4]

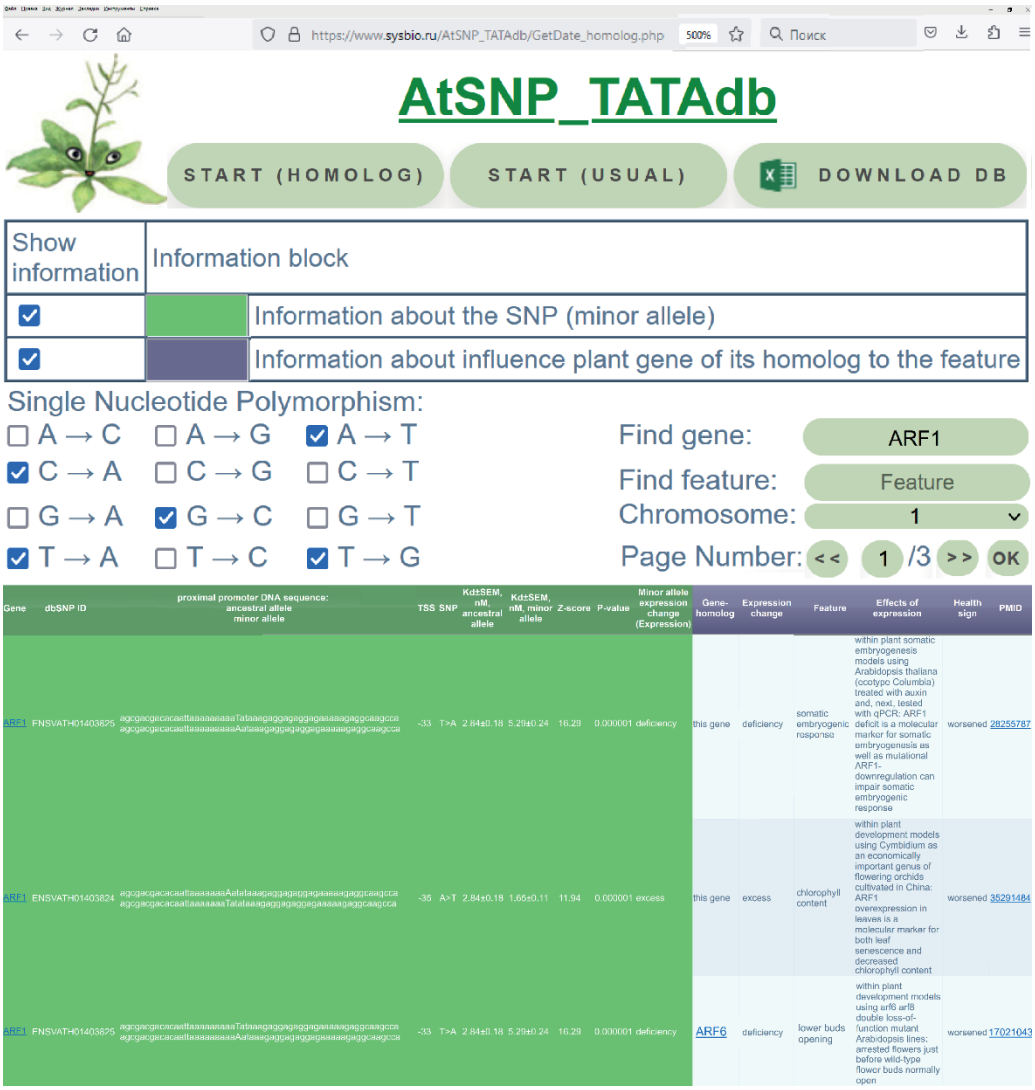


Fig. 2. A sample entry in the AtSNP_TATAdb documents [5] on two SNPs ENSVATH01403825:A and ENSVATH01403824:T of the *ARF1* (*Auxin Response Factor 1)* gene, which can significantly downregulate and upregulate its expression, as calculated in this work together with their annotation using the database PubMed [4] (PMIDs: 28255787, 35291484, 17021043)

References

1. Rasskazov D., Chadaeva I., Sharypova E. et al. Plant_SNP_TATA_Z-Tester: A Web service that unequivocally estimates the impact of proximal promoter mutations on plant gene expression. *Int J Mol Sci*. 2022;23(15):8684. doi 10.3390/ijms23158684
2. Contreras-Moreira B., Naamati G., Rosello M. et al. Scripting analyses of genomes in Ensembl Plants. *Methods Mol Biol*. 2022;2443:27-55. doi 10.1007/978-1-0716-2067-0_2
3. Berardini T.Z., Reiser L., Li D., Mezheritsky Y., Muller R., Strait E., Huala E. The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis*. 2015;53(8):474-485. doi 10.1002/dvg.22877
4. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)*. 2011;2011:baq036. doi 10.1093/database/baq036
5. Bogomolov A., Zolotareva K., Filonov S. et al. AtSNP_TATAdb: candidate molecular markers of plant advantages related to single nucleotide polymorphisms within proximal promoters of Arabidopsis thaliana L. *Int J Mol Sci*. 2024;25(1):607. doi 10.3390/ijms25010607

2 Симпозиум «Системная компьютерная биология»

Symposium “Systems computational biology”

doi 10.18699/bgrs2024-2.1-01

Реконструкция и анализ молекулярно-генетических механизмов регуляции метаболических путей при глиобластоме с использованием данных ВЭЖХ-МС/МС

Адамовская А.В.[1, 2]*, Басов Н.В.[2, 3], Деменков П.С.[1, 2], Гайслер Е.В.[2], Рогачев А.Д.[2, 3], Иванисенко В.А.[1, 2], Мишинов С.В.[4], Ступак В.В.[4], Чересиз С.В.[2], Олешко О.С.[2], Покровский А.Г.[2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия*

[4] *Новосибирский научно-исследовательский институт травматологии и ортопедии им. Я.Л. Цивьяна, Новосибирск, Россия*

** a.adamovskaya@g.nsu.ru*

Ключевые слова: глиобластома; ВЭЖХ-МС/МС; метаболомный скрининг; генные сети; ANDSytem

Мотивация и цель: Глиобластома – наиболее распространенная злокачественная опухоль головного мозга, характеризующаяся низкой пятилетней выживаемостью и высокой частотой рецидивов [1]. Известно, что клетки претерпевают метаболические трансформации в процессе онкогенеза, а метаболизм в опухолях отличается от нормального метаболизма. Изучение метаболома опухоли головного мозга, а также поиск диагностических и прогностических биомаркеров имеют ключевое значение для улучшения выживаемости пациентов и разработки новых персонализированных методов лечения. Цель данной работы – реконструкция и анализ молекулярно-генетических механизмов регуляции метаболических путей, нарушенных при глиобластоме, с использованием данных метаболомного анализа.

Методы и алгоритмы: Методом ВЭЖХ-МС/МС, разработанным нами ранее [2], был выполнен таргетированный скрининг метаболитов тканей глиобластомы и перитуморального пространства. Для определения значимости различий между значениями уровня метаболитов в глиобластоме и перитуморальной зоне использовали комбинацию критерия Манна–Уитни и теста Cucconi. Перепредставленность метаболических путей KEGG оценивалась с помощью анализа обогащенности, представленного в веб-инструменте MetaboAnalyst 6.0. Реконструкция и анализ генных сетей проводились с использованием программно-информационной системы ANDSystem [3].

Результаты: По результатам статистической обработки данных метаболомного скрининга, содержание ряда метаболитов различалось в тканях глиобластомы и перитуморального пространства. Анализ перепредставленности метаболических путей KEGG для найденного набора значимых метаболитов, выполненный в веб-инструменте MetaboAnalyst 6.0, позволил выявить, что метаболизм сфинголипидов является статистически значимым метаболическим процессом (FDR < 0.05). С помощью программно-информационной системы ANDSystem были реконструированы и проанализированы генные сети, описывающие генетическую регуляцию

ферментов метаболизма сфинголипидов с участием генетических маркеров глиобластомы. Было показано, что ферменты SPHK1, NEUR1 и ASM подвержены воздействию наибольшего количества белков, ассоциированных с развитием и прогрессированием глиобластомы. При этом наибольшее число ферментов регулируется маркерами P53 и TNFA.
Выводы: В настоящей работе с использованием статистических тестов были идентифицированы потенциальные метаболические маркеры глиобластомы на основе анализа метаболомных данных тканей опухоли и перитуморального пространства. Анализ перепредставленности метаболических путей и реконструкция регуляторных генных сетей позволили выдвинуть гипотезы о связи нарушений функционирования метаболизма сфинголипидов с известными генетическими маркерами глиобластомы, включая P53, EGFR, HIF2α и другие.
Финансирование: Исследование поддержано Министерством науки и высшего образования РФ (государственное задание No. FSUS-2020-0035).

Reconstruction and analysis of molecular-genetic mechanisms of metabolic pathways regulation in glioblastoma using LC-MS/MS data

Adamovskaya A.V.[1, 2]*, Basov N.V.[2, 3], Demenkov P.S.[1, 2], Gaisler E.V.[2], Rogachev A.D.[2, 3], Ivanisenko V.A.[1, 2], Mishinov S.V.[4], Stupak V.V.[4], Cheresiz S.V.[2], Oleshko O.S.[2], Pokrovsky A.G.[2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia*

[4] *FSBI "Novosibirsk Research Institute of Traumatology and Orthopedics named after Ya.L. Tsiviyan", Novosibirsk, Russia*

** a.adamovskaya@g.nsu.ru*

Key words: glioblastoma; LC-MS/MS; metabolomic screening; gene network; ANDSystem

Motivation and Aim: Glioblastoma is the most common malignant brain tumor, characterized by low five-year survival and high relapse rates [1]. Malignant cells, including brain tumors, are known to exhibit metabolic changes. Studying the brain tumor metabolome, as well as searching for diagnostic and prognostic biomarkers, is key to improving patient survival and developing new personalized treatments. The purpose of this work is to reconstruct and analyze the molecular genetic mechanisms of regulation of metabolic pathways disrupted in glioblastoma using metabolomic analysis data.
Methods and Algorithms: Using the LC-MS/MS method, which we previously developed [2], we performed targeted screening of metabolites in glioblastoma tissues and the peritumoral brain zone. To determine the significance of differences between the values of metabolite levels in glioblastoma and the peritumoral zone, a combination of the Mann-Whitney test and the Cucconi test was used. The overrepresentation of KEGG metabolic pathways was assessed using enrichment analysis provided in the MetaboAnalyst 6.0 web tool. Reconstruction and analysis of gene networks were carried out using the ANDSystem [3].

Results: According to the results of statistical processing of metabolomic screening data, the content of a number of metabolites differed in the tissues of glioblastoma and peritumoral space. Analysis of the overrepresentation of KEGG metabolic pathways for the found set of significant metabolites, performed in the MetaboAnalyst 6.0 web tool, revealed that sphingolipid metabolism is a statistically significant metabolic process (FDR < 0.05). Using the ANDSystem, gene networks describing the genetic regulation of enzymes of sphingolipid metabolism involving genetic markers of glioblastoma were reconstructed and analyzed. The enzymes SPHK1, NEUR1 and ASM have been shown to be affected by the largest number of proteins associated with the development and progression of glioblastoma. In this case, the largest number of enzymes is regulated by the P53 and TNFA markers.
Conclusion: In this work, using statistical tests, potential metabolic markers of glioblastoma were identified based on the analysis of metabolomic data from tumor tissues and peritumoral space. Analysis of the overrepresentation of metabolic pathways and reconstruction of regulatory gene networks allowed us to put forward hypotheses about the connection between disorders of sphingolipid metabolism and known genetic markers of glioblastoma, including P53, EGFR, HIF2α and others.
Funding: The study was supported by the Ministry of Science and Higher Education of the Russian Federation (state project No. FSUS-2020-0035).

Список литературы/References

1. Omuro A., DeAngelis L.M. Glioblastoma and other malignant gliomas: a clinical review. *JAMA*. 2013;310(17):1842-1850
2. Basov N.V., Rogachev A.D. et al. Global LC-MS/MS targeted metabolomics using a combination of HILIC and RP LC separation modes on an organic monolithic column based on 1-vinyl-1, 2, 4- triazole. *Talanta*. 2024;267:125168
3. Ivanisenko V.A. et al. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl 1):34

doi 10.18699/bgrs2024-2.1-02

Реконструкция и статистические оценки генных сетей комплексных заболеваний на основе онлайн-инструментов биоинформатики

Волков И.А.[1], Анашкина А.А.[1, 2], Туркина В.А.[1], Савина Е.А.[1, 2]*, Орлов Ю.Л.[1, 3]*

[1] *Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия*

[2] *Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

[3] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

** kaparos@list.ru, orlov@bionet.nsc.ru*

Ключевые слова: медицинская информатика; медицинская генетика; онкология; комплексные заболевания; генные онтологии; генные сети; образование

Мотивация и цель: Исследование генных сетей комплексных заболеваний является важной задачей биомедицины, требующей интеграции данных. Бурный рост данных технологий секвенирования следующего поколения (NGS, next generation sequencing) требует разработки новых методов анализа, процессинга и предварительной обработки данных секвенирования с использованием математических и статистических методов, включая реализацию программ для клиник, для врачей, с удобным интерфейсом. Несмотря на существующий широкий набор компьютерных программ в мире, для решения конкретных задач анализа клинических данных в России необходима их адаптация, в том числе создание биоинформационных методик, учебных пособий и обучающих элементов для студентов, интернов и ординаторов медицинских учреждений [1, 2]. Отработка методики реконструкции генных сетей важна как для поиска генов-мишеней для терапии, так и для развития биоинформационных методов, поиска новых учебных решений для образования. Представлен научный проект исследования комплексных заболеваний, в которых сложно выделить генетические компоненты, включая в том числе ментальные расстройства, шизофрению, болезнь Паркинсона [3]. Практические цели проекта – разработка специализированных компьютерных программных модулей для реконструкции и анализа генных сетей комплексных заболеваний, которые могут быть использованы для задач цифрового образования и повышения квалификации, подготовки студентов биомедицинских и биоинженерных специальностей.

Методы и алгоритмы: Собраны имеющиеся программные инструменты, подготовлен конвейер обработки данных по созданию списков генов. Составление списков генов выполнялось на основе запросов к базам данных OMIM (Online Mendelian Inheritance in Man), GeneCards, MalaCards. Список уточнялся с использованием данных о дифференциальной экспрессии генов, включая некодирующие РНК (из базы данных TCGA), и публикаций. Использовалась база данных геномов рака GDC Data Portal. Расчет генных онтологий производился с помощью открытых ресурсов для биоинформационного анализа: PANTHER (http://pantherdb.org/) и DAVID (https://davidbioinformatics.nih.gov/), для визуализации генных онтологий использовался ресурс g:GOSt (http://biit.cs.ut.ee/gprofiler/gost). Анализ ключевых

генов проводился с помощью базы данных GeneCards.org и каталога соматических мутаций при раке COSMIC (https://cancer.sanger.ac.uk/cosmic/).
Компьютерное исследование данных секвенирования выполняется на основе интеграции доступных данных секвенирования (RNA-seq и ChIP-seq из GEO NCBI) и компьютерных ресурсов: ArrayExpress, TCGA, ENCODE (ENCyclopedia Of DNA Elements), UCSC Genome Browser, а также российских компьютерных ресурсов TRRD, GeneNet, NPRD, SitEx и комплексов ICGenomics, ANDSystem. Использование данных секвенирования, представленных в GEO NCBI, выполнялось с помощью ресурса GEO Dataset Browser.
Результаты: В качестве основных примеров приложений рассмотрены задачи анализа опухолей мозга – для глиомы, менингиомы, с исследованием осложнений и сопутствующих заболеваний, связанных с инфекциями, в том числе по данным после пандемии коронавирусной инфекции, которые накоплены за последние годы. В качестве дополнительных научных приложений с использованием той же методики реконструкции функциональных сетей генов, белков и метаболитов выполнена компьютерная реконструкция генных сетей для ряда комплексных заболеваний: онкологических заболеваний – глиомы, рака молочной железы, колоректального рака, и ряда ментальных расстройств, таких как болезнь Паркинсона и ментальных расстройств [4]. При этом отработана методика анализа сети, оценки ее структуры, устойчивости к воздействию лекарственных соединений на гены-мишени, будут выполнены статистические оценки структуры генных сетей, оценки распределения числа связей, участия некодирующих РНК [5].
Представлен онлайн-ресурс для интеграции данных по ассоциированным генам комплексных заболеваний.
Выводы: Представлена методика биоинформационного исследования генов глиомы. Собран список ключевых генов заболевания, рассчитаны категории генных онтологий, построена сеть белок-белковых взаимодействий, выделены наиболее крупные кластеры генов. Рассмотрены структуры белков и потенциал белок-белковых взаимодействий для узловых генов рассмотренных генных сетей, таких как TP53, PTEN.
Финансирование: Исследование поддержано РНФ (грант 24-24-00563, «Разработка цифровых образовательных программ в биомедицине»).

Reconstruction and statistical estimates of gene networks of complex diseases based on online bioinformatics tools

Volkov I.A.[1], Anashkina A.A.[1, 2], Turkina V.A.[1], Savina E.A.[1, 2]*, Orlov Y.L.[1, 3]*
[1] *Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*
[2] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*
[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
* *kaparos@list.ru, orlov@bionet.nsc.ru*

Key words: medical informatics; medical genetics; oncology; complex diseases; gene ontologies; gene networks; education

Motivation and Aim: The study of gene networks of complex diseases is an important biomedical task that requires data integration. The rapid growth of these next generation

sequencing technologies (NGS) requires the development of new methods for analyzing, processing and preprocessing sequencing data using mathematical and statistical methods, including the implementation of programs for clinics, for doctors, with a user-friendly interface. Despite the existing wide range of computer programs in the world, in order to solve specific tasks of analyzing clinical data in Russia, their adaptation is necessary, including the creation of bioinformatic techniques, textbooks and training elements for students, interns and residents of medical institutions [1, 2]. The development of methods for the reconstruction of gene networks is important both for the search for target genes for therapy, and the development of bioinformatic methods, the search for new educational solutions for the education of students based on available online bioinformatics tools. A scientific project for the study of complex diseases in which it is difficult to identify genetic components, including mental disorders, schizophrenia, and Parkinson's disease, is presented [3]. The practical goals of the project are the development of specialized computer software modules for the reconstruction and analysis of gene networks of complex diseases, which can be used for digital education and advanced training, training students of biomedical and bioengineering specialties, digital transformation of education.

Methods and Algorithms: The available software tools have been collected, a data processing pipeline for creating gene lists has been prepared. The list of genes was compiled based on current queries to the OMIM (Online Mendelian Inheritance in Man), GeneCards, and MalaCards databases. The list was refined using data on the differential expression of genes, including non-coding RNAs (from the TCGA database), and publications. The GDC Cancer genome database Data Portal was used. The calculation of gene ontologies was performed using open resources for bioinformatic analysis: PANTHER (http://pantherdb.org/) and DAVID (https://davidbioinformatics.nih.gov/), the g:GOSt resource was used to visualize gene ontologies (http://biit.cs.ut.ee/gprofiler/gost). The analysis of key genes was carried out using a database GeneCards.org and the catalog of somatic mutations in cancer COSMIC (https://cancer.sanger.ac.uk/cosmic/).

The computer study of sequencing data is performed based on the integration of available sequencing data (RNA-seq and ChIP-seq from GEO NCBI) and computer resources: ArrayExpress, TCGA, ENCODE (ENCyclopedia Of DNA Elements), UCSC Genome Browser, as well as Russian computer resources TRRD, GeneNet, NPRD, SitEx and ICGenomics, ANDSystem complexes. The sequencing data presented in GEO NCBI was used using the GEO Dataset Browser resource.

Results: As the main examples of applications, the tasks of analyzing brain tumors are considered – for glioma, meningioma, with the study of complications and concomitant diseases associated with infections, including according to data after the coronavirus pandemic, which have accumulated in recent years. As additional scientific applications, using the same technique of reconstruction of functional networks of genes, proteins and metabolites, computer reconstruction of gene networks for a number of complex diseases – oncological diseases – glioma, breast cancer, colorectal cancer, and a number of mental disorders such as Parkinson's disease and metal disorders was performed [4].

At the same time, a methodology for analyzing the network, evaluating its structure, resistance to the effects of medicinal compounds on target genes has been developed, statistical estimates of the structure of gene networks, estimates of the distribution of the number of connections, and the participation of non-coding RNAs will be performed [5]. An online resource for integrating data on associated genes of complex diseases is presented.

Conclusion: The technique of bioinformatics research of glioma genes is presented. A list of key genes of the disease has been compiled, categories of gene ontologies have been calculated, a network of protein-protein interactions has been built, and the largest clusters of genes have been identified. The structures of proteins and the potential of protein-protein interactions for the nodal genes of the considered gene networks, such as TP53, PTEN, are considered.

Funding: The study was supported by the Russian Science Foundation (grant 24-24-00563).

Список литературы/References

1. Turkina V.A., Orlova N.G., Orlov Y.L. Biophysics education section and computational training discussion at VII Congress of Russian Biophysicists. *Biophys Rev*. 2023;15:807-809. doi 10.1007/s12551-023-01147-5
2. Anashkina A.A., Rubin A.B., Gudimchuk N.B. et al. VII Congress of Russian Biophysicists – 2023, Krasnodar, Russia. *Biophys Rev*. 2023;15:801-805. doi 10.1007/s12551-023-01164-4
3. Орлов Ю.Л., Галиева А.Г., Орлова Н.Г., Иванова Е.Н., Мозылева Ю.А., Анашкина А.А. Реконструкция генной сети болезни Паркинсона для поиска генов-мишеней. *Биомедицинская химия*. 2021;67(3):222-230. doi 10.18097/pbmc20216703222
   [Orlov Y.L., Galieva A.G., Orlova N.G. et al. Reconstruction of gene network associated with Parkinson disease for gene targets search. *Biomeditsinskaia Khimiia*. 2021;67(3):222-230. doi 10.18097/pbmc20216703222 (in Russian)]
4. Voropaeva E.N., Pospelova T.I., Orlov Y.L, Churkina M.I., Berezina O.V., Gurazheva A.A., Ageeva T.A., Seregina O.B., Maksimov V.N. The Methylation of the p53 Targets the Genes MIR-203, MIR-129-2, MIR-34A and MIR-34B/C in the Tumor Tissue of Diffuse Large B-Cell Lymphoma. *Genes*. 2022;13(8):1401. doi 10.3390/genes13081401
5. Dergilev A., Orlova N., Dobrovolskaya O., Orlov Y. Statistical estimates of multiple transcription factors binding in the model plant genomes based on ChIP-seq data. *J Integr Bioinf*. 2022;19(1):20200036. doi 10.1515/jib-2020-0036

doi 10.18699/bgrs2024-2.1-03

Идентификация и описание ключевых генов, связанных с патогенезом расстройств аутистического спектра посредством реконструкции и анализа генных сетей

Колос А.В.[1]*, Можина В.В.[2], Куцин И.Ю.[3], Минасазова А.Р.[1], Ефимочкина С.М.[1], Савина Е.А.[1]

[1] *Первый Московский государственный медицинский университет им. И.М. Сеченова, Москва, Россия*

[2] *ФГБОУ ВО МЗ РФ, Волгоградский государственный медицинский университет, Волгоград, Россия*

[3] *ФГБОУ ВО МЗ РФ, Омский государственный медицинский университет, Омск, Россия*

* *kolos_a_v@student.sechenov.ru*

Ключевые слова: расстройства аутистического спектра; генные сети; анализ генных онтологий; глутаматергические синапсы

Мотивация и цель: На сегодняшний день точные молекулярные механизмы, лежащие за биологией расстройств аутистического спектра (РАС), не выяснены. Однако количество «надежных» генов, выявленных разными исследовательскими группами, продолжает увеличиваться [1]. Несмотря на гетерогенную природу РАС, определенные гены встречаются одновременно в различных списках, вызывая интерес к пониманию их роли и взаимодействий друг с другом. В этом контексте представляется важным проанализировать эти гены и их взаимосвязи с использованием системного подхода, такого как реконструкция генных сетей.

Методы и алгоритмы: Для получения списка ключевых генов использовалась база данных ассоциации генов с нарушениями нервного развития GeneTrek [2], внутри которой были выбраны списки, содержащие гены, ассоциированные с расстройствами аутистического спектра, а именно: SFARI [3] категории 1 – high-confidence, SPARK [4], SysNDD [5] и DDG2P [6] категории definitive genes, DBD с мутациями, приводящими к потере функции [7]. Реконструкция генной сети была выполнена в STRINGdb, затем экспортирована в Cystoscape для дальнейшего анализа и визуализации генных онтологий. Анализ генных онтологий для списка генов был выполнен с помощью PANTHER Overrepresentation Test [8]. Использовался тест Фишера с поправкой Бонферрони на множественное тестирование. Получившийся список из 71 гена аннотировали с помощью SynGo v1.2 – аннотации синаптических генных онтологий [9].

Результаты: Генная сеть связная и образует функциональные кластеры, на что указывают значения PPI enrichment *p*-value генной сети < 1.0e–16 и средний коэффициент локальной кластеризации, равный 0.493. Средняя степень узла равна 9.44, а количество ребер – 335, из которых 240 получены путем текстового майнинга научных публикаций, 26 – из результатов экспериментов, 28 – из баз данных и 9 из них – данные коэкспрессии. Большинство коэкспрессируемых пар участвуют в регуляции апоптоза, клеточной пролиферации и нейрогенезе. Согласно SynGO, среди 71 гена 21 ген имеет синаптическую локализацию и/или функцию. В то время как анализ генных онтологий молекулярной функции

(GO molecular function) в PANTHER определил как значимые онтологии: активность протеин-цистеинметилтрансферазы с *p*-value = 3.70E–02, среди которых S-аденозилметионин-зависимая активность метилтрансферазы, занимающаяся переносом одноуглеродных групп с *p*-value = 3.95E–02, и активность протеинметилтрансферазы с *p*-value = 1.45E–04, активность ацетилтрансферазы гистона H3K27 с *p*-value = 3.70E–02, неметилированное связывание CpG посредством связывания органических циклических соединений с *p*-value = 4.81E–03 и АТФ-зависимая активность ремоделирования хроматина 5.56E–04. А анализ оверрепрезентации генных онтологий клеточной локализации выявил как наиболее значимые: пресинаптическую мембрану с *p*-value = 3.13E–03, хромосомную и центромерную область с *p*-value = 3.47E-03, глутаматергические синапсы *p*-value = 2.73E–02 и хроматин с *p*-value = 1.91E–08. Что позволяет предполагать, что глутаматергическая гипотеза патофизиологии РАС [10] может стоять за значительным количеством фенотипов РАС.

Выводы: Реконструкция генной сети генов, обнаруживаемых независимыми исследовательскими группами, позволила определить наличие функциональных кластеров РАС. Учитывая высокую гетерогенность, вероятно, что на настоящий момент гены, экспрессирующиеся в пресинаптической мембране, хромосомной и центромерной области, а также глутаматергических синапсах, превалируют в большом количестве фенотипов аутизма. Это представляется важным для развития возможных путей рациональной терапии, которые не представлялись возможными до эпохи успешного открытия генов.

Финансирование: Исследование поддержано РНФ (грант 24-24-00563, «Разработка цифровых образовательных программ в биомедицине»).

Identification and description of key genes associated with the pathogenesis of autism spectrum disorders through reconstruction and analysis of gene networks

Kolos A.V.[1]*, Mozhina V.V. [2], Kutsin I.Yu [3], Minasazova A.R.[1], Efimochkina S.M.[1], Savina E.A.[1]

[1] *I.M. Sechenov First Moscow State Medical University, Moscow, Russia*

[2] *Volgograd State Medical University, Volgograd, Russia*

[3] *Omsk State Medical University, Omsk, Russia*

* *kolos_a_v@student.sechenov.ru*

Key words: autism spectrum disorder; gene networks; gene ontology analysis; glutamatergic synapses

Motivation and Aim: To date, the precise molecular mechanisms underlying the biology of autism spectrum disorders (ASD) remain unclear. However, the number of “high-confidence” genes identified by various research groups continues to grow [1]. Despite the heterogeneous nature of ASD, certain genes appear concurrently in different lists, sparking interest in understanding their roles and interactions with each other. In this context, it is important to analyze these genes and their relationships using a systems approach, such as gene network reconstruction.

Methods and Algorithms: To obtain a list of key genes, the GeneTrek database linking genes to neurodevelopmental disorders was utilized [2], from which lists were selected

containing genes associated with autism spectrum disorders, specifically SFARI [3] category 1 – high-confidence, SPARK [4], SysNDD [5], and DDG2P [6] definitive genes categories, as well as DBD with loss-of-function mutations [7]. Gene network reconstruction was conducted in STRINGdb, then exported to Cytoscape for further analysis and visualization of gene ontologies. Analysis of gene ontologies for the gene list was performed using the PANTHER Overrepresentation Test [8]. The Fisher test with Bonferroni correction for multiple testing was employed. The resulting list of 71 genes was annotated using SynGo v1.2 – annotations of synaptic gene ontologies [9].

Results: The gene network is cohesive and forms functional clusters, as indicated by the PPI enrichment p-value of the gene network < 1.0e-16 and an average clustering coefficient of 0.493. The average node degree is 9.44, with 335 edges, of which 240 are derived from text mining of scientific publications, 26 from experimental findings, 28 from databases, and 9 from co-expression data. Most co-expressed gene pairs are involved in the regulation of apoptosis, cell proliferation, and neurogenesis. According to SynGO, among the 71 genes, 21 genes have synaptic localization and/or function. Meanwhile, analysis of gene ontologies in molecular function (GO molecular function) using PANTHER identified significant ontologies: protein-cysteine S-methyltransferase activity with a p-value of 3.70E-02, including S-adenosylmethionine-dependent methyltransferase activity involved in one-carbon group transfer with a p-value of 3.95E-02 and protein methyltransferase activity with a p-value of 1.45E-04, histone H3K27 acetyltransferase activity with a p-value of 3.70E-02, unmethylated CpG binding through binding of organic cyclic compounds with a p-value of 4.81E-03, and ATP-dependent chromatin remodeling activity at 5.56E-04. Overrepresentation analysis of gene ontologies in cellular localization identified the most significant as: presynaptic membrane with a p-value of 3.13E-03, chromosomal and centromeric regions with a p-value of 3.47E-03, glutamatergic synapses with a p-value of 2.73E-02, and chromatin with a p-value of 1.91E-08. This suggests that the glutamatergic hypothesis of ASD pathophysiology [10] may underlie a significant number of ASD phenotypes.

Conclusion: Reconstruction of a gene network of genes identified by independent research groups has allowed for the identification of functional clusters related to autism spectrum disorders (ASD). Given the high heterogeneity, it is likely that currently, genes expressed in the presynaptic membrane, chromosomal and centromeric regions, as well as glutamatergic synapses, prevail in a large number of autism phenotypes. This is crucial for the development of potential pathways for rational therapy that were not feasible before the era of successful gene discovery.

Funding: This work was supported by Russian Science Foundation (grant project 24-24-00563, “Development of digital education programs in biomedicine”).

Список литературы/References

1. Manoli D.S., State M.W. Autism Spectrum Disorder Genetics and the Search for Pathological Mechanisms. *Am J Psychiatry*. 2021;178(1):30-38. doi 10.1176/appi.ajp.2020.20111608
2. Leblond C.S., Le T.L., Malesys S., Cliquet F., Tabet A.C., Delorme R., Rolland T., Bourgeron T. Operative list of genes associated with autism and neurodevelopmental disorders based on database review. *Mol Cell Neurosci*. 2021;113:103623. doi 10.1016/j.mcn.2021.103623
3. Abrahams B.S., Arking D.E., Campbell D.B. et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism*. 2013;4:36. doi 10.1186/2040-2392-4-36
4. SPARK Consortium. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron*. 2018;97(3):488-493. doi 10.1016/j.neuron.2018.01.015
5. Kochinke K., Zweier C., Nijhof B. et al. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *Am J Hum Genet.* 2016;98(1):149-164. doi 10.1016/j.ajhg.2015.11.024

6. Wright C.F., Fitzgerald T.W., Jones W.D., Clayton S., McRae J.F. et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet*. 2015;385(9975):1305-1314. doi 10.1016/S0140-6736(14)61705-0
7. Gonzalez-Mantilla A.J., Moreno-De-Luca A. Ledbetter D.H., Martin C.L. A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. *JAMA Psychiatry*. 2016;73(3):275-283. doi 10.1001/jamapsychiatry.2015.2692
8. Mi H., Muruganujan A., Ebert D., Huang X., Thomas P.D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*. 2019;47(D1):D419-D426. doi 10.1093/nar/gky1038
9. Koopmans F., van Nierop P., Andres-Alonso M., Byrnes A., Cijsouw T., Coba M.P., Cornelisse L.N. et al. SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron*. 2019;103(2):217-234.e4. doi 10.1016/j.neuron.2019.05.002
10. Galineau L., Arlicot N., Dupont A.C., Briend F., Houy-Durand E. et al. Glutamatergic synapse in autism: a complex story for a complex disorder. *Mol Psychiatry*. 2023;28(2):801-809. doi 10.1038/s41380-022-01860-9

doi 10.18699/bgrs2024-2.1-04

Анализ корреляций в протеомах микроорганизмов

Лазарева А.*, Матюшкина Д., Бутенко И., Говорун В.
Институт системной биологии и медицины Роспотребнадзора, Москва, Россия
** a.lazareva@sysbiomed.ru*

Ключевые слова: протеомика; системная биология; синтетическая биология

Мотивация и цель: История изучения клетки стартовала еще в XVII веке с рассмотрения Робертом Гуком структуры пробки в микроскоп. С тех пор научные методы претерпели множество изменений, которые позволили изучить все детали: от строения отдельной молекулы до принципиальных связей в каскадах метаболических реакций; от концентраций ферментов и метаболитов в клетке до их точной локализации. Однако современной науке так и не удалось приблизиться к полному пониманию и описанию всех процессов жизнедеятельности клетки.
К настоящему моменту научным сообществом накоплен большой объем «мультиомиксных» данных для различных биологических объектов, из которых исследователи используют только «верхушку айсберга» для проверки конкретных гипотез, при этом гигантский объем информации остается неиспользованным. В свою очередь, тщательный их анализ позволит выявить новые закономерности и механизмы организации и регуляции живой клетки. Эти знания необходимы для возможности создания искусственной упрощенной клетки, которая обеспечит прорыв как в фундаментальной науке о жизни, так и в прикладных исследованиях. И в первую очередь для реализации данного подхода в нашей работе была поставлена следующая цель: выявление минимального белкового «кора» живой клетки и анализ закономерностей его организации и функционирования.
Методы и алгоритмы: Культивирование *Mycoplasma gallisepticum S6*. Лизис клеток. Трипсинолиз в растворе. Количественный протеомный анализ методами MRM и shot-gun. Расчет z-оценки. Расчет коэффициента корреляции методом Пирсона. Анализ методом главных компонент (PCA). Кластеризация методами t-SNE, K-средних и DBSCAN.
Результаты: В системном подходе к изучению микроорганизмов есть несколько ключевых этапов, в том числе секвенирование генома и сопутствующего генетического материала, аннотация генома, установление механизмов регуляции элементов клетки, аннотация метаболических путей и различных биохимических процессов и т. д. Существуют как общепризнанные модельные микроорганизмы, для которых получен большой набор мультиомиксных данных [1–3], так и достаточно хорошо изученные минимальные бактериальные клетки [4, 5]. В ходе работы были отобраны протеомные данные для *Mycoplasma gallisepticum S6*, *Escherichia coli*, *Bacillus subtilis* и *Saccharomyces cerevisiae* в различных экспериментальных условиях, оценено необходимое значение количества образцов, после которого коэффициент корреляции переставал изменяться [1–4]. В зависимости от экспериментальных выборок (точнее степени их гомогенности) требовалось от 20 до 50 образцов для достижения ошибки определения коэффициента корреляции < 0.5 %. Далее были построены взвешенные графы, где в вершинах располагались

белки, а ребрами отмечены корреляционные связи, при этом вес ребра была обратно пропорционально связана с модулем коэффициента корреляции. Для удобства анализа скоррелированных и антискоррелированных пар ребра были покрашены в различные цвета (красный – для положительного коэффициента корреляции, синий – для отрицательного). Еще на этом этапе было обнаружено двудольное распределение микоплазменного протеома, где две группы плотно скоррелированных белков были антискоррелированы между собой. Повышение модуля значений коэффициентов корреляции у отображаемых ребер с 0.75 до 0.85 привело к исчезновению связей между кластерами, что позволяет сделать вывод о более плотном взаимодействии внутри обнаруженных групп.

При этом интересным результатом оказалось то, что первый кластер сформировали в основном белки, обладающие различной ферментативной активностью, а во втором кластере обнаружилось много белков, взаимодействующих с нуклеиновыми кислотами. При этом наложение метаболических путей или данных о белковых комплексах не позволило объяснить подобную кластеризацию. Поэтому полученные результаты дают возможность предположить, что белки в кластерах организуются по определенным правилам, а не случайным образом. Возможно, такая протеомная кластеризация необходима живой клетке для поддержания ее метастабильного состояния.

Для определения кластеризации протеома микоплазмы другим способом, с помощью метода главных компонент, были определены основные белки, участвующие в формировании реакции. Для микоплазмы оказалось достаточно всего 15 белков, чтобы обеспечить 70 % ответа на внешний стимул. Снизив с помощью метода t-SNE размерность белкового пространства до двумерного, все белки были отображены на координатной плоскости, а затем были размечены по результатам кластеризации с помощью методов k-means и DBSCAN. С помощью первого обнаружился оптимальный диапазон числа кластеров: от 3 до 5. Метод DBSCAN разделил протеом на 4 группы. Однако эти результаты не удалось соотнести с результатами корреляционного анализа, оставляя вопрос группировки белков открытым. Аналогичное исследование протеомов *Escherichia coli* и *Bacillus subtilis* также показало двудольную корреляционную организацию, образованную плотными связями внутри групп и более слабыми между. Однако усложнение организации и регуляции относительно *Mycoplasma gallisepticum* привело к увеличению модулей коэффициентов корреляций между белками. В случае же эукариотического протеома *Saccharomyces cerevisiae* подобная двудольная организация была утрачена.

Выводы: Были отобраны и подготовлены протеомные данные в различных условиях для *Mycoplasma gallisepticum S6* (данные, полученные нами), *Escherichia coli*, *Bacillus subtilis* и *Saccharomyces cerevisiae* (результаты из баз данных).

Были применены методы кластеризации протеомных данных с помощью алгоритмов корреляционного анализа, а также с помощью PCA, k-means, t-SNE, DBSCAN, а также проанализированы результаты корреляционного анализа и выявлены общие закономерности. Для полученных кластеров характерна двудольная организация в протеомах прокариот (*Mycoplasma gallisepticum*, *Escherichia coli*, *Bacillus subtilis*) и отсутствие подобного разделения для эукариот (*Saccharomyces cerevisiae*). А методы кластеризации k-means и DBSCAN на данных для *Mycoplasma gallisepticum* показали существование групп белков со схожим поведением при изменении условий существования клетки.

Для *Mycoplasma gallisepticum S6* были дополнительно подготовлены протеомные данные для культур в нормальных и осмотических условиях. В каждом из наборов данных проведен корреляционный анализ, показавший динамику изменения соотношений в протеомах при стрессовых воздействиях.
Финансирование: Исследование поддержано Управлением Роспотребнадзора ГЗ № 1022040800170-3-1.6.23 «Создание искусственных клеточных систем».

Analysis of correlations between microbial proteomes

Lazareva A.*, Matyushkina D., Butenko I., Govorun V.
Scientific Research Institute for Systems Biology and Medicine, Moscow, Russia
** a.lazareva@sysbiomed.ru*

Key words: proteomics; systems biology; synthetic biology

Motivation and Aim: The history of cell research started back in the XVII century with Robert Hooke's examination of the structure of a cell in a microscope. Since then, scientific methods have undergone many changes that have made it possible to study all the details: from the structure of a single molecule to the fundamental connections in the cascades of metabolic reactions; from the concentrations of enzymes and metabolites in the cell to their exact localization. However, modern science has not been able to come close to a complete understanding and description of all the processes of cell activity.
Up to the present time, the scientific community has accumulated a significant amount of "multiomics" data on various biological objects, of which researchers have only utilized a small fraction to test specific hypotheses, while a huge amount of remains untapped. A thorough analysis of this data could reveal novel patterns and mechanisms of organization and regulation of living cells. This knowledge would be essential for the development of an artificial simplified cell that could lead to a breakthrough in both fundamental life sciences and applied research. And first, in order to implement this approach, we set the following objective in our work: to identify the minimum protein "core" of a living cell and to analyze the patterns of their organization and function.
Methods and Algorithms: Cultivation of *Mycoplasma gallisepticum S6* cells. Cell lysis. Trypsinolysis in solution. Quantitative proteomic analysis using MRM and shotgun methods. Calculation of the Z-score. Pearson correlation analysis. PCA (Principal Component Analysis). TSN (T-SNE) (t-distributed stochastic neighbor embedding). K-means clustering. DBSCAN (Density-based spatial clustering of applications with noise).
Results: As previously mentioned, there are several crucial stages in the examination of an organism. These include sequencing the genome and associated genetic material, establishing a connection between a gene and protein, identifying the functions and methods of regulating cellular components, annotating metabolic pathways and diverse biochemical processes, among others. There are both widely recognized model microorganisms, for which a significant volume of multiomics data has been collected [1–3], and well-researched minimal bacterial cells [4, 5]. During the course of this work, proteomic data was collected for *Mycoplasma gallisepticum S6*, *Escherichia coli*, *Bacillus subtilis*, and *Saccharomyces cerevisiae* under various experimental conditions. The required number of samples was then estimated, following which the correlation coefficient stopped changing [1–4]. Depending on the experimental samples and the degree of their

homogeneity, it took between 20 and 50 samples to achieve an error of less than 0.5 % in determining the correlation coefficient. Next, weighted graphs were created, with proteins as vertices and correlation connections represented by edges. The edge length was inversely proportional to the absolute value of the correlation coefficient. To facilitate analysis of correlated and anticorrelated pairs, edges were color-coded, with red indicating a positive correlation and blue indicating a negative correlation. Even at this stage, a bipartite distribution of the Mycoplasma proteome was observed, where two groups of highly correlated proteins were anticorrelated with each other. An increase in the magnitude of the correlation coefficient for the observed edges from 0.75 to 0.85 resulted in the disappearance of connections between clusters, indicating a denser interaction within the identified groups.

Interestingly, the first cluster consisted primarily of proteins with diverse enzymatic activities, while the second cluster contained proteins interacting with nucleic acids. The overlap of metabolic pathways or information on protein complexes was insufficient to explain this clustering. Consequently, the findings suggest that proteins within clusters are organized in a manner consistent with specific rules, rather than randomly. It is possible that such proteomic clustering plays a role in maintaining the metastable state of a living cell.

To identify the clustering of mycoplasma proteomes in another manner, using the principal components analysis, the key proteins involved in reaction formation were identified. For mycoplasmas, only 15 proteins were sufficient to account for 70 % of the response to external stimuli. By reducing the protein space dimension to two dimensions through the t-SNE method, all proteins were plotted on a two-dimensional coordinate plane and then categorized according to the clustering results obtained from the k-means and DBSCAN methods. Using the first one, the optimal number of clusters range was found to be between 3 and 5. However, the DBSCAN approach divided the proteome into four groups, which could not be correlated with correlation analysis results, leaving the issue of protein grouping unresolved.

A similar study of the proteomic profiles of *Escherichia coli* and *Bacillus subtilis* also revealed a bipartite organization, characterized by strong intra-group interactions and weaker inter-group connections. However, due to the complexity of the regulatory mechanisms and organization in *Mycoplasma gallisepticum*, there has been an increase in the number of modules of correlation coefficients between proteins compared to other organisms. In contrast, the eukaryotic proteome of *Saccharomyces cerevisiae* exhibits a more complex organization and lacks the bipartite structure observed in other organisms.

Conclusion: Proteomic data were collected and processed under various conditions for *Mycoplasma gallisepticum S6* (data generated by our team), *Escherichia coli*, *Bacillus subtilis*, and *Saccharomyces cerevisiae* (data retrieved from publicly available databases).

Methods of clustering proteomic data were applied using correlation analysis algorithms as well as PCA, k-means, t-SNE and DBSCAN. The results of correlation analysis were used for identifying general patterns. The clusters obtained were characterized by a bipartite organization in prokaryotic proteomes, such as *Mycoplasma gallisepticum*, *Escherichia coli* and *Bacillus subtilis*, and the absence of such separation in eukaryotic proteomes like *Saccharomyces cerevisiae*. Additionally, clustering methods such as k-means, t-SNE and DBSCAN on data for *Mycoplasma gallisepticum* showed the existence of groups of proteins that exhibit similar behavior under varying conditions of cellular existence.

Proteomic data for cultures under normal and osmotic conditions was additionally prepared for *Mycoplasma gallisepticum S6*. For each data set, a correlation analysis was conducted, which revealed the dynamics of changes in proteome ratios under stress conditions.

Funding: The study is supported by Rospotrebnadzor service state task No. 1022040800170-3-1.6.23.

Список литературы/References

1. Mori M. et al. From coarse to fine: the absolute *Escherichia coli* proteome under diverse growth conditions. *Mol Syst Biol*. 2021;17(5):e9536. doi 10.15252/msb.20209536
2. Shi L. et al. Tyrosine 601 of Bacillus subtilis DnaK Undergoes Phosphorylation and Is Crucial for Chaperone Activity and Heat Shock Survival. *Front Microbiol.* 2016;7;533. doi 10.3389/fmicb.2016.00533
3. Ho B. et al. Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. *Cell Syst*. 2018;6(2):192-205. doi 10.1016/j.cels.2017.12.004
4. Butenko I. et al. Data-independent proteome profile of Mycoplasma gallisepticum under normal conditions and heat stress. *Data Brief.* 2017;16:700-704. doi 10.1016/j.dib.2017.11.093
5. Hutchison C. et al. Design and synthesis of a minimal bacterial genome. *Science*. 2016;351(6280):aad6253. doi 10.1126/science.aad6253

doi 10.18699/bgrs2024-2.1-05

Генная сеть клеточного ответа на терагерцовое излучение, реконструированная на основе анализа метаболомных данных

Макарова А.А.[1]*, Бутикова Е.А.[3], Басов Н.В.[2, 3], Рогачев А.Д.[2, 3], Иванисенко Т.В.[1, 4], Деменков П.С.[1, 3, 4], Гайслер Е.В.[3], Покровский А.Г.[3], Колчанов Н.А.[1, 3, 4], Иванисенко В.А.[1, 3, 4]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия*

[3] *Новосибирский государственный университет, Новосибирск, Россия*

[4] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

* *MakarovaAA@bionet.nsc.ru*

Ключевые слова: генные сети; ТГц; терагерцовое излучение; метаболомный скрининг

Мотивация и цель: В связи с развитием технологий применения терагерцового (ТГц) излучения необходимо исследовать его воздействие на живые организмы. Известно, что в культурах клеток, подвергшихся ТГц излучению, наблюдаются нарушения ряда биологических процессов, ассоциированные с изменением экспрессионных [1–3] и протеомных [4] профилей. В частности, описана дисрегуляция биологических процессов, протекающих в митохондриях [1]. Мы предположили, что воздействие ТГц излучения на мембраны митохондрий может приводить к повреждению супрамолекулярных мембранных структур, в частности липидных рафтов, нарушая регуляцию функций белков транспорта и превращения метаболитов. В рамках исследования этой гипотезы целью работы являлась интерпретация метаболомных данных с применением подхода генных сетей.

Методы и алгоритмы: Для анализа были взяты данные LC-MS/MS метаболомного скрининга клеток меланомы человека, проведенного через трое суток после облучения лазером с частотой 2.3 ТГц и для контрольной группы клеток – термостатирования в течение трех суток. Статистический анализ данных проводился с применением критерия Манна–Уитни с поправкой на множественное тестирование Benjamini–Hochberg.

Для построения генной сети из баз данных KEGG и HMDB были извлечены ферменты биосинтеза и деградации выявленных метаболитов; митохондриальные мембранные белки человека экстрагированы из базы данных MitoProteome. Реконструкция генной сети производилась в программе ANDVisio – графическом интерфейсе программно-информационной системы ANDSystem [5]. Для анализа перепредставленности биологических процессов и клеточных компонентов Gene Ontology был использован веб-инструмент DAVID.

Результаты: Сравнительный статистический анализ метаболомных профилей группы клеток, подвергшихся ТГц излучению, и контрольной группы показал значимые различия для 40 метаболитов, среди которых оказались пурины, пиримидины и другие связанные с энергетическими процессами метаболиты. На основа-

нии нашей гипотезы о влиянии ТГц излучения на структуры в митохондриальной мембране была реконструирована генная сеть регуляции ферментов превращения значимых метаболитов митохондриальными белками-транспортерами (рис. 1).

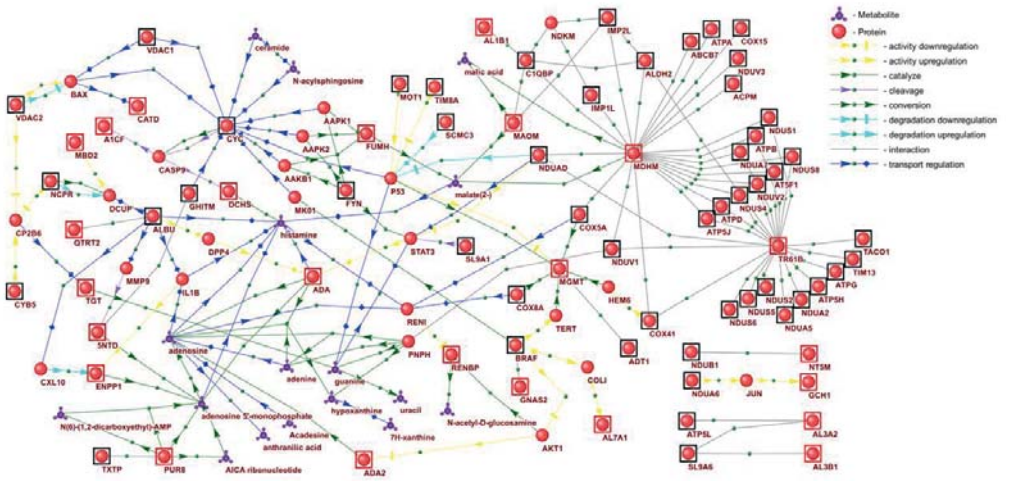


Рис. 1. Генная сеть регуляции ферментов биосинтеза и деградации значимых метаболитов. Черными рамками отмечены митохондриальные мембранные белки, красными рамками – ферменты превращения метаболитов, белки без рамок – промежуточные участники регуляторных путей

В генной сети выявлены регуляторные пути, потенциально нарушенные в клетке вследствие ТГц излучения. В качестве важных участников генной сети мы отметили ферменты MDHM, TR61B, MGMT, митохондриальные АТФ-синтазы и NADH-дегидрогеназы. Перепредставленные биологические процессы Gene Ontology включали метаболические процессы в митохондриях, цепь переноса электронов, регуляцию апоптоза и ответ клетки на внешний стимул. Анализ перепредставленности клеточных компонентов Gene Ontology идентифицировал в генной сети 8 белков липидных рафтов, указывая на роль рафтоподобных структур в регуляции превращения значимых метаболитов и их транспорта между цитоплазмой и митохондрией.

Выводы: Анализ метаболомных профилей клеток меланомы человека, подвергшихся ТГц излучению, идентифицировал 40 значимо измененных метаболитов. Реконструирована генная сеть регуляции ферментов биосинтеза и деградации метаболитов митохондриальными мембранными белками. Среди участников генной сети идентифицированы белки липидных рафтов. Из анализа генной сети выявлены механизмы, потенциально нарушенные в клетках, подвергшихся ТГц излучению. Согласно нашей гипотезе, структура липидных рафтов нарушается вследствие воздействия ТГц излучения, что способствует изменению транспорта и реакций превращения метаболитов.

Финансирование: Биоинформатический анализ был проведен при поддержке бюджетного проекта FWNR-2022-0020.

The gene network of cellular response to Terahertz radiation based on the metabolomic data analysis

Makarova A.A.[1]*, Butikova E.A.[3], Basov N.V.[2, 3], Rogachev A.D.[2, 3], Ivanisenko T.V.[1, 4], Demenkov P.S.[1, 3, 4], Gaisler E.V.[3], Pokrovsky A.G.[3], Kolchanov N.A.[1, 3, 4], Ivanisenko V.A.[1, 3, 4]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

[4] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *MakarovaAA@bionet.nsc.ru*

Key words: gene networks; THz; terahertz irradiation; metabolomic screening

Motivation and Aim: In connection with the development of terahertz (THz) radiation applicative technologies, it is necessary to investigate its effects on living organisms. Studies based on transcriptomic [1–3] and proteomic [4] analysis reported that THz radiation may disturb a number of biological processes, in particular, the metabolic processes in mitochondria [1]. We hypothesized that THz radiation acts upon lipid raft-like microdomains on mitochondrial membranes, leading to the disrupted regulation of transport and conversion of metabolites. In the course of our hypothesis, we aimed to analyze the metabolomic profiles of the cells exposed to THz radiation with the usage of gene networks approach.

Methods and Algorithms: The data taken into analysis included the results of LC-MS/MS metabolomic screening performed on human melanoma cells three days after irradiation with 2.3 THz laser, and the cells in control group were thermostated for three days. Statistical analysis of the data was performed using the Mann–Whitney U test with Benjamini–Hochberg correction. For the gene network reconstruction, the enzymes of biosynthesis and degradation of significantly altered metabolites were extracted from KEGG and HMDB databases; the mitochondrial membrane proteins were extracted from MitoProteome database. The gene network reconstruction was performed with ANDVisio program – the user interface of the ANDSystem software and information system [5]. The web tool DAVID was used for overrepresentation analysis of Gene Ontology biological processes and cellular components.

Results: The comparative statistical analysis of metabolomic profiles of the group of THz irradiated cells and the control group revealed significant differences for 40 metabolites, including purines, pyrimidines, and other energy-related metabolites. Based on our hypothesis about the effect of THz radiation on the mitochondrial membrane structures, we reconstructed the gene network of regulation of the enzymes involved in biosynthesis and degradation of significant metabolites (Fig. 1). Regulatory pathways to the enzymes started from mitochondrial membrane proteins implementing the transport function.

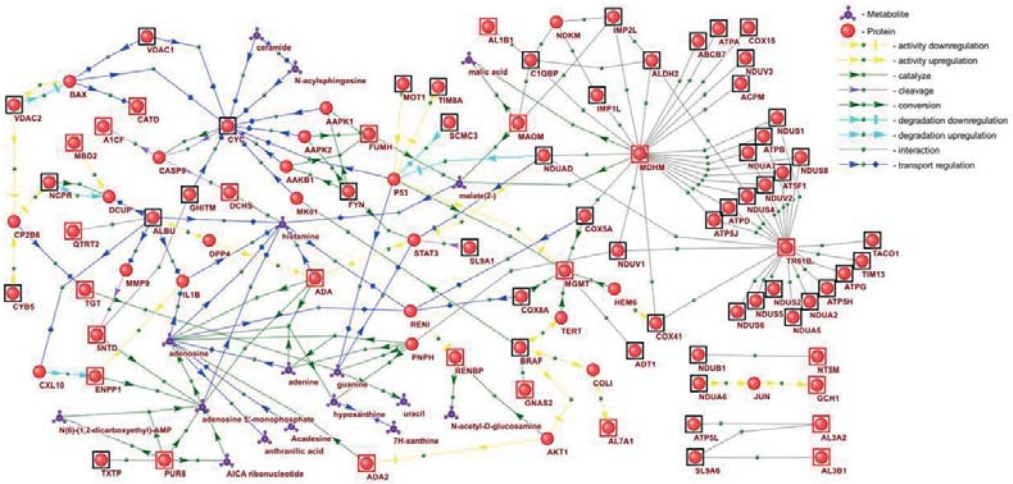


Fig. 1. Gene network of regulation of the enzymes involved in conversion of significant metabolites. Red-framed spheres represent enzymes of significant metabolites conversion. Black-framed spheres – mitochondrial membrane proteins. Red-spheres without frames – intermediate proteins

The gene network identified regulatory pathways potentially disrupted in the cell subsequent to THz-irradiation. We noted the following gene network participants: the enzymes MDHM, TR61B, MGMT, mitochondrial ATP synthases, and NADH dehydrogenases as the important components of the pathways. Overrepresented Gene Ontology biological processes included mitochondrial metabolic processes, electron transport chain, regulation of apoptosis and cell response to external stimulus. The analysis of overrepresented Gene Ontology cellular components identified 8 proteins in the gene network annotated in membrane rafts, indicating the role of raft-like structures in regulating conversion and transport of significant metabolites.

Conclusion: The analysis of metabolic profiles of THz-irradiated human melanoma cells identified 40 significantly altered metabolites. We reconstructed the gene network of regulation of the enzymes involved in biosynthesis and degradation of significant metabolites. Lipid raft proteins were identified among the participants of the gene network. The gene network analysis revealed the mechanisms potentially disrupted in THz-irradiated cells. According to our hypothesis, the structure of lipid rafts is disturbed due to THz exposure, resulting in alterations in catalytic reactions and transport of metabolites.

Funding: Bioinformatic analysis was supported by the budget project (No. FWNR-2022-0020).

Список литературы/References

1. Tachizaki T. et al. Terahertz pulse-altered gene networks in human induced pluripotent stem cells. *Optics Lett*. 2020;45(21):6078-6081
2. Bogomazova A.N. et al. No DNA damage response and negligible genome-wide transcriptional changes in human embryonic stem cells exposed to terahertz radiation. *Sci Rep*. 2015;5(1):7749
3. Alexandrov B.S. et al. Specificity and heterogeneity of terahertz radiation effect on gene expression in mouse mesenchymal stem cells. *Sci Rep*. 2013;3(1):1184
4. Bannikova S. et al. Specific Features of the Proteomic Response of Thermophilic Bacterium *Geobacillus icigianus* to Terahertz Irradiation. *Int J Mol Sci*. 2022;23(23):15216
5. Ivanisenko V.A. et al. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst Biol*. 2015;9:1-10

doi 10.18699/bgrs2024-2.1-06

Реконструкция генных сетей комплексных ментальных расстройств

Минасазова А.Р.[1], Братчикова М.С.[1], Волков И.А.[1], Савина Е.А.[1, 2]*, Орлов Ю.Л.[1, 3]*

[1] *Первый МГМУ им. И.М.Сеченова Минздрава России (Сеченовский Университет), Москва, Россия*

[2] *Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

[3] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

** kaparos@list.ru, orlov@bionet.nsc.ru*

Ключевые слова: медицинская информатика; медицинская генетика; онкология; комплексные заболевания; генные онтологии; генные сети; образование

Мотивация и цель: Исследование генных сетей комплексных заболеваний является важной задачей биомедицины, требующей интеграции данных.

Представлен научный проект исследования комплексных заболеваний, в которых сложно выделить генетические компоненты, в том числе ментальные расстройства, шизофрению, болезнь Паркинсона [1]. Шизофрения – тяжелое полиморфное расстройство психики (или группа психических расстройств), для которого характерен распад процессов мышления и эмоциональных реакций [2, 3]. По данным Global Health Data Exchange (GHDx) за 2019 г., шизофренией страдают примерно 0.32 % человек во всем мире, среди взрослых этот показатель составляет 0.45 %. Шизофрения является плохо изученным заболеванием с многообразием симптомов, характерных для иных патологических состояний, и сложной диагностикой без однозначного лечения. Активный исследовательский интерес к данной патологии обусловлен ее неуклонным ростом: распространение заболевания выросло почти в два раза, с 13.1 миллиона (1990 г.) до 20.9 миллионов (2016 г.) [4]. Для поиска генов-мишеней терапии необходима реконструкция генной сети шизофрении – комплекса взаимодействующих макромолекул, функционально связанных с заболеванием. Дальнейший анализ предполагает кластеризацию генов в сети, выявление ключевых генов, обладающих наибольшим числом контактов в сети, нахождение категорий генных онтологий.

Методы и алгоритмы: Собраны имеющиеся программные инструменты, подготовлен конвейер обработки данных по созданию списков генов. Составление списков генов выполнялось на основе запросов к базам данных OMIM (Online Mendelian Inheritance in Man), GeneCards, MalaCards. Анализ генных онтологий для списка генов был выполнен с помощью ресурса PANTHER (Protein ANalysis THrough Evolutionary Relationships) (http://pantherdb.org/). Для реконструкции сети использовались инструменты STRING-DB, GeneMANIA, приложения Cytosacape. С помощью онлайн-инструментов биоинформатики OMIM, PANTHER и DAVID, GeneMANIA и STRING-DB, GeneCards был проанализирован актуальный массив данных, связанных с шизофренией, рассчитаны категории генных онтологий для списка из 200 генов, включая биологические процессы, молекулярные функции, и клеточные компоненты.

Результаты: В качестве основных примеров приложений рассмотрены задачи анализа генных сетей ряда ментальных расстройств, таких как шизофрения, болезнь Паркинсона [1]. При этом отработана методика анализа сети, оценки ее структуры и кластеризации [5]. Значимые категории генных онтологий отражают влияние шизофрении на передачу нейронных импульсов, экспрессию генов в тканях мозга. Визуализированы и построены генные сети, содержащие выявленные ключевые объекты и их взаимосвязи, позволяющие выявить взаимодействия генов; выделен сильно связанный кластер, включающий гены BDNF, SLC6A4, HTR2A, HTR2C, CHRM1, SRC, AKT, YWHAE, DISC1, DRD2, COMT, NDEL1, NOS1, CAMK28.

При анализе реконструированной генной сети можно сделать вывод, что наибольшее количество связей имеет ген DISC1. Мутации, возникающие в данном гене, связаны как с шизофренией, так и с другими психическими расстройствами, например, биполярным расстройством. Также можно выделить несколько кластеров, самый большой из которых включает гены BDNF, SLC6A4, HTR2A, HTR2C. В целом представленные гены участвуют в регуляции передачи синаптических сигналов и в передаче межклеточных сигналов при помощи нейромедиаторов. Генами с самой большой оценкой релевантности для шизофрении являются COMT, DISC1, HTR2A, NRXN1.

С использованием инструментов биоинформатики описаны функции этих генов. COMT – катехол-О-метилтрансфераза – катализирует перенос метильной группы с S-аденозилметионина на катехоламины, включая нейротрансмиттеры допамин, эпинефрин и норэпинефрин. Это О-метилирование приводит к одному из основных путей деградации катехоламиновых трансмиттеров. DISC1 участвует в регуляции многочисленных аспектов эмбрионального и взрослого нейрогенеза [6]. Ген необходим для пролиферации нейронных предшественников в вентрикулярной/субвентрикулярной зоне во время эмбрионального развития мозга и в зубчатой извилине гиппокампа у взрослых. Участвует в Wnt-опосредованной пролиферации нейронных предшественников в качестве положительного регулятора, играет роль модулятора сигнального пути AKT-mTOR, контролирующего темп процесса интеграции новорожденных нейронов в ходе взрослого нейрогенеза, включая позиционирование нейронов, развитие дендритов и формирование синапсов. HTR2A кодирует один из рецепторов для серотонина. Мутации в этом гене связаны с предрасположенностью к шизофрении и обсессивно-компульсивному расстройству, а также с ответом на антидепрессант циталопрам у пациентов с большим депрессивным расстройством (БДР). У пациентов с БДР, имеющих мутацию в интроне 2 этого гена, значительно снижен ответ на циталопрам, поскольку данный антидепрессант снижает экспрессию гена. NRXN1 кодирует однопроходный мембранный белок типа I, принадлежащий к семейству нейрексинов. Нейрексины – это рецепторы клеточной поверхности, которые связывают нейролигины, комплексы нейрексин/нейролигин в синапсах центральной нервной системы. Этот комплекс необходим для эффективной нейротрансмиссии и участвует в формировании синаптических контактов. Мутации в этом гене связаны с синдромом Питта-Хопкинса-2 и могут способствовать предрасположенности к шизофрении.

Выводы: Представлена методика биоинформационного исследования генных сетей ментальных расстройств. На примере шизофрении определены наиболее релевантные гены. Анализ генов, связанных с шизофренией, определение их положения в генной сети (связанности) позволяют установить взаимосвязи, установить,

какие гены и их продукты являются ключевыми при протекании болезни, оценить их перспективность в качестве генов-мишеней для лекарственных воздействий. Практическое использование полученных результатов остается сложной задачей, так как шизофрения является генетически комплексной болезнью с большой расходимостью причин и условий возникновения.
Финансирование: Исследование поддержано РНФ (грант 24-24-00563, «Разработка цифровых образовательных программ в биомедицине»).

Reconstruction of gene networks of complex mental disorders

Minasazova A.R.[1], Bratchikova M.S.[1], Volkov I.A.[1], Savina E.A.[1, 2]*, Orlov Y.L.[1, 3]*
[1] *Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*
[2] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*
[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
* *kaparos@list.ru, orlov@bionet.nsc.ru*

Key words: medical informatics; medical genetics; oncology; complex diseases; gene ontologies; gene networks; education

Motivation and Aim: The study of gene networks of complex diseases is an important biomedical task that requires data integration.
A scientific project investigating complex diseases in which it is difficult to identify genetic components is presented, including mental disorders, schizophrenia, and Parkinson's disease [1]. Schizophrenia is a severe polymorphic mental disorder (or group of mental disorders) characterised by a breakdown of thought processes and emotional reactions [2, 3]. According to the Global Health Data Exchange (GHDx) 2019 data, schizophrenia affects approximately 0.32 % of people worldwide; among adults, the prevalence is 0.45 %. Schizophrenia is a poorly understood disorder with a variety of symptoms characteristic of other pathological conditions and a difficult diagnosis without unequivocal treatment. The active research interest in this pathology is due to its steady increase (the prevalence of the disease has almost doubled, from 13.1 million (1990) to 20.9 million (2016) [4]). In order to search for target genes for therapy, it is necessary to reconstruct the schizophrenia gene network – a complex of interacting macromolecules functionally related to the disease. Further analysis involves clustering the genes in the network, identifying the key genes with the highest number of contacts in the network, and finding categories of gene ontologies.
Methods and Algorithms: Available software tools were assembled, and a data processing pipeline for creating gene lists was prepared. The list of genes was compiled based on current queries to the OMIM (Online Mendelian Inheritance in Man), GeneCards, and MalaCards databases. Gene ontologies for the gene list were analysed using the PANTHER (Protein ANalysis THrough Evolutionary Relationships) resource (http://pantherdb.org/). STRING-DB, GeneMANIA, and Cytosacape tools were used for network reconstruction. Using the online bioinformatics tools OMIM, PANTHER and DAVID, GeneMANIA and STRING-DB, GeneCards analysed an up-to-date dataset related to schizophrenia, calculating gene ontology categories for a list of 200 genes, including biological processes, molecular functions, and cellular components.

Results: As the main examples of applications are the analysis of gene networks of a number of mental disorders, such as schizophrenia and Parkinson's disease [1]. At the same time, the methodology of network analysis, evaluation of its structure, and clustering has been worked out [5]. Significant categories of gene ontologies reflect the impact of schizophrenia on neural impulse transmission, gene expression in brain tissue. Gene networks containing the identified key sites and their interconnections were visualised and constructed to reveal gene interactions; a highly connected cluster including the genes BDNF, SLC6A4, HTR2A, HTR2C, CHRM1, SRC, AKT, YWHAE, DISC1, DRD2, COMT, NDEL1, NOS1, CAMK28 was identified.

When analysing the reconstructed gene network, it can be concluded that the DISC1 gene has the highest number of connections. Mutations occurring in this gene are associated with both schizophrenia and other mental disorders, such as bipolar disorder. Several clusters can also be identified, the largest of which includes the genes BDNF, SLC6A4, HTR2A, HTR2C. In general, the presented genes are involved in the regulation of synaptic signal transmission and in the transmission of intercellular signals by means of neurotransmitters. The genes with the highest relevance score for schizophrenia are COMT, DISC1, HTR2A, NRXN1.

Using bioinformatics tools, the functions of these genes are described. COMT – Catechol-O-methyltransferase catalyses the transfer of a methyl group from S-adenosylmethionine to catecholamines, including the neurotransmitters dopamine, epinephrine and norepinephrine. This O-methylation leads to one of the major degradation pathways of catecholamine transmitters. DISC1 is involved in the regulation of numerous aspects of embryonic and adult neurogenesis [6]. The gene is required for proliferation of neural precursors in the ventricular/subventricular zone during embryonic brain development and in the dentate gyrus of the hippocampus in adults. It participates in Wnt-mediated proliferation of neural progenitors as a positive regulator, and plays a role as a modulator of the AKT-mTOR signalling pathway that controls the rate of integration of newborn neurons during adult neurogenesis, including neuronal positioning, dendrite development and synapse formation. HTR2A encodes one of the receptors for serotonin. Mutations in this gene are associated with predisposition to schizophrenia and obsessive-compulsive disorder, as well as response to the antidepressant citalopram in patients with major depressive disorder (MDD). MDD patients with a mutation in intron 2 of this gene have a significantly reduced response to citalopram because this antidepressant reduces the expression of this gene. NRXN1 encodes a type I single-pass membrane protein belonging to the neurexin family. Neurexins are cell surface receptors that bind neuroligins, neurexin/neuroligin complexes at synapses in the central nervous system. This complex is required for efficient neurotransmission and is involved in the formation of synaptic contacts. Mutations in this gene are associated with Pitt-Hopkins syndrome-2 and may contribute to a predisposition to schizophrenia.

Conclusion: The technique of bioinformatics study of gene networks of mental disorders is presented. Using schizophrenia as an example, the most relevant genes are identified. The analysis of genes related to schizophrenia, determination of their position in the gene network (connectivity) allows to establish interrelationships, to determine which genes and their products are key in the course of the disease, to assess their prospects as target genes for drug interventions. The practical use of the results obtained remains a difficult task, since schizophrenia is a genetically complex disease with a large divergence of causes and conditions of occurrence.

Funding: The study was supported by the Russian Science Foundation (grant 24-24-00563).

Список литературы/References

1. Орлов Ю.Л., Галиева А.Г., Орлова Н.Г., Иванова Е.Н., Мозылева Ю.А., Анашкина А.А. Реконструкция генной сети болезни Паркинсона для поиска генов-мишеней. *Биомед. химия*. 2021;67(3):222-230
   [Orlov Y.L., Galieva A.G., Orlova N.G. et al. Reconstruction of gene network associated with Parkinson disease for gene targets search. *Biomeditsinskaia Khimiia*. 2021;67(3):222-230. doi 10.18097/pbmc20216703222 (in Russian)]
2. Charlson F.J., Ferrari A.J., Santomauro D.F., Diminic S., Stockings E., Scott J.G., McGrath J.J., Whiteford H.A. Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016. *Schizophr Bull*. 2018;44(6):1195-1203. doi 10.1093/schbul/sby058
3. Дохоян А.Ю., Глущенко М.В., Орлов Ю.Л. Реконструкция генной сети шизофрении для поиска генов-мишеней. *Ульяновский мед.-биол. журн*. 2022;3:6-22. doi 10.34014/2227-1848-2022-3-6-22
   [Dokhoyan A.Y., Glushchenko M.V., Orlov Y.L. Reconstruction of schizophrenia gene network in search for target genes. *Ul'yanovskiy Mediko-Biologicheskiy Zhurnal*. 2022;3:6-22. doi 10.34014/2227-1848-2022-3-6-22 (in Russian)]
4. Owen M.J., Sawa A., Mortensen P.B. Schizophrenia. *Lancet*. 2016;388(10039):86-97. doi 10.1016/S0140-6736(15)01121-6
5. Turkina V.A., Orlova N.G., Orlov Y.L. Biophysics education section and computational training discussion at VII Congress of Russian Biophysicists. *Biophys. Rev*. 2023;15:807-809. doi 10.1007/s12551-023-01147-5
6. Mao Y., Ge X., Frank C.L., Madison J.M., Koehler A.N., Doud M.K., Tassa C., Berry E.M., Soda T., Singh K.K., Biechele T., Petryshen T.L., Moon R.T., Haggarty S.J., Tsai L.H. Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3beta/beta-catenin signaling. *Cell*. 2009;136(6):1017-31. doi 10.1016/j.cell.2008.12.044

doi 10.18699/bgrs2024-2.1-07

Модель энергетического баланса в клетках с учетом аномальной активности ретротранспозонов, характерной для раковых клеток

Павлов С.Р.[1], Гурский В.В.[1, 2]*, Самсонова М.Г.[1], Канапин А.А.[1, 3], Самсонова А.А.[1, 3]

[1] *Лаборатория математической биологии и биоинформатики, Санкт-Петербургский политехнический университет Петра Великого (СПбПУ), Санкт-Петербург, Россия*

[2] *Теоретический отдел, Физико-технический институт им. А.Ф. Иоффе РАН, Санкт-Петербург, Россия*

[3] *Центр Вычислительной Биологии, Санкт-Петербургский политехнический университет Петра Великого (СПбПУ), Санкт-Петербург, Россия*

* *gursky@math.ioffe.ru*

Ключевые слова: ретротранспозоны; мобильные генетические элементы, противораковая терапия; апоптоз, математическое моделирование; энергетический баланс

Мотивация и цель: Активность ретротранспозонов в клетках повышена при многих видах рака, при этом развитие опухолей и других патологических процессов может быть связано с нарушением механизмов регуляции ретротранспозонов [1]. В то же время исследования методов борьбы с раком на модельных животных показывают важную роль ретротранспозонов в активации противоопухолевых защитных механизмов [2]. Представленная работа направлена на исследованиее возможности использования ретротранспозонов для запуска механизмов клеточной смерти в результате нарушения клеточного энергетического баланса и снижения общего количества доступных молекул АТФ в клетке. Известно, что даже относительно небольшое снижение уровня АТФ запускает необратимые процессы, приводящие к некрозу или апоптозу клетки [3]. Одним из факторов, приводящим к снижению количества свободных молекул АТФ, может служить повышенный расход энергии на процессы, связанные с аномальным распространением ретротранспозонов в геноме. Для борьбы с поврежденными и раковыми клетками и запуска апоптоза представляется целесообразным воздействовать на механизмы регуляции ретротранспозонов, оказывающие наибольшее влияние на уровень АТФ в клетке.

Методы и алгоритмы: Для исследования энергетического состояния клетки разработана математическая модель, которая отражает динамическое распределение энергозатрат на основные клеточные процессы, включая расходы на активность ретротранспозонов. Полученная модель представляет собой систему обыкновенных дифференциальных уравнений, описывающих изменение концентраций АТФ, активных ретротранспозонов в геноме, а также мРНК и белков, являющихся продуктами экспрессии ретротранспозонов и конститутивных генов. Уравнения модели сформулированы с помощью методов кинетического моделирования и включают в себя скорости расхода энергии на основные клеточные процессы: репликация ДНК, транскрипция и трансляция, интеграция новых копий ретротранс-

позонов в геном и процессы, связанные с деградацией имеющихся в клетке молекул и деактивацией ретротранспозонов вследствие действия защитных клеточных механизмов. В модели рассматривается два вида ретротранспозонов, активных на сегодняшний день в геноме человека, – LINE-1 и SINE. Взаимодействие этих ретротранспозонов между собой математически представимо в виде классических уравнений типа «хищник-жертва», где LINE-1 производит белок ORF2p, содержащий обратную транскриптазу для встройки новых копий ретротранспозонов в геном, и является «жертвой», поскольку SINE использует этот белок для своих встроек и тем самым является «хищником». В общем виде модель представляет собой систему обобщенных уравнений Лотки–Вольтерры и отражает три вида конкуренции: конкуренцию за АТФ со стороны ретротранспозонов и конститутивных генов, конкуренцию за рибосомы со стороны мРНК LINE-1 и мРНК конститутивных генов и конкуренцию за ORF2p между LINE-1 и SINE. Значения большей части параметров в модели оценены по литературным данным [4]. Остальная часть параметров подгонялась таким образом, чтобы получить стационарное решение, в котором значения динамических переменных соответствовали бы оценкам, имеющимся в литературе. Компьютерная реализация модели выполнена на языке Python.

Результаты: Полученная модель имеет единственное устойчивое стационарное решение, при котором и LINE-1 и SINE сохраняют низкую активность. Предполагается, что данный сценарий реализуется в клетке в «нормальных условиях» и представляет собой референсное решение. Показано существование двух других стационарных решений, соответствующих случаям деактивации SINE в геноме и полной деактивации ретротранспозонов и их продуктов в клетке. Режим, при котором SINE становится неактивным, характеризуется трехкратным снижением количества АТФ в стационарном состоянии из-за повышенного потребления энергии на процессы, связанные с распространением LINE-1.

Для нахождения параметров, изменение которых оказывает наибольшее влияние на уровень АТФ в клетке, произведен анализ чувствительности параметров. Для параметров с наибольшей чувствительностью, связанных с ретротранспозонами, были вычислены новые значения динамических переменных в стационаре при изменении величины параметра. Наибольшая прямая зависимость количества доступных АТФ от параметров была выявлена для константы скорости деактивации LINE-1, а наибольшая обратная зависимость – для максимальной скорости транскрипции LINE-1. Моделирование показало, что взаимное изменение этих параметров приводит к кумулятивному эффекту, и для трехкратного снижения уровня АТФ достаточно изменения этих параметров в три раза относительно референсных значений. Того же эффекта можно достичь при раздельном пятикратном изменении указанных параметров.

Анализ распределения энергозатрат на различные процессы в клетке показал, что в случае возмущения параметров с целью снижения уровня АТФ основной расход АТФ значительно смещается с трансляции конститутивных генов в референсном решении на транскрипцию ретротранспозонов.

Модель позволила исследовать динамические эффекты изменения количества доступных молекул АТФ, связанные с начальной концентрацией ретротранспозонов в клетке. Трехкратное кратковременное уменьшение уровня АТФ в клетке возможно при увеличении начального количества LINE-1 в 100 раз. При увеличении LINE-1 в 1000 раз снижение количества доступных молекул АТФ в клетке более

чем в три раза длится в течение 13 минут, а максимальное снижение АТФ будет более чем десятикратным.
Выводы: Разработанная математическая модель отражает энергетический баланс в клетке в условиях конкуренции за клеточные ресурсы со стороны активных ретротранспозонов типа LINE-1 и SINE. В результате исследования в рамках модели сценариев перераспределения энергозатрат внутри клетки при нарушении механизмов регуляции транспозонов показана возможность значительного влияния на клеточные ресурсы со стороны транспозонов при модификации их скорости транскрипции и при дерегуляции механизмов сайленсинга. Описанный эффект уменьшения количества свободных молекул АТФ в клетке при возмущении параметров в модели может быть использован для инициации программ клеточной гибели раковых клеток. Результаты моделирования свидетельствуют о принципиальной возможности реализации сценариев уничтожения опухолевых клеток за счет увеличения нагрузки на энергетический баланс со стороны активизирующихся транспозонов.
Финансирование: Работа выполнена при финансовой поддержке Российского научного фонда, грант № 23-24-00153.

Model of energy balance in cells influenced by the cancer-specific abnormal activity of retrotransposons

Pavlov S.R.[1], Gursky V.V.[1, 2]*, Samsonova M.G.[1], Kanapin A.A.[1, 3], Samsonova A.A.[1, 3]

[1] *Mathematical Biology and Bioinformatics Laboratory, Peter the Great Saint Petersburg Polytechnic University, St. Petersburg, Russia*

[2] *Theoretical Department, Ioffe Institute, RAS, St. Petersburg, Russia*

[3] *Center for Computational Biology, Peter the Great Saint Petersburg Polytechnic University, St. Petersburg, Russia*

* *gursky@math.ioffe.ru*

Key words: retrotransposons; mobile genetic elements; anticancer therapy; apoptosis, mathematical modeling; energy balance

Motivation and aim: The retrotransposon activity in cells is increased in many types of cancer, and the tumor development and other pathological processes may be associated with disruption of the regulatory mechanisms of retrotransposons [1]. At the same time, studies of cancer control methods in animal models show the important role of retrotransposons in the activation of antitumor defense mechanisms [2]. The presented work is aimed at studying the possibility of using retrotransposons to trigger cell death mechanisms as a result of disruption of the cellular energy balance and a decrease in the total amount of available ATP molecules in the cell. It is known that even a relatively small decrease in ATP levels triggers irreversible processes leading to cell necrosis or apoptosis [3]. One of the factors leading to a decrease in the amount of free ATP molecules may be the increased energy consumption by processes associated with the abnormal activity of retrotransposons in the genome. Therefore, to trigger apoptosis in damaged and cancer cells, it seems reasonable to alter the regulatory mechanisms of retrotransposons that have the greatest impact on the ATP level in the cell.

Methods and algorithms: We have modeled the energy state of the cell by developing a mathematical model that reflects the dynamic distribution of energy costs for basic cellular processes, including costs for the retrotransposon activity. The resulting model is a system of ordinary differential equations that describe changes in the concentrations of ATP, active retrotransposons in the genome, as well as mRNA and proteins that are expression products of retrotransposons and constitutive genes. The model equations are formulated using kinetic modeling methods and include the energy consumption rates for basic cellular processes: DNA replication, transcription and translation, retrotransposon integration into the genome, and processes associated with the degradation of molecules present in the cell and the retrotransposon deactivation due to the action of cellular protection mechanisms. The model considers two types of retrotransposons that are currently active in the human genome: LINE-1 and SINE. The mutual interaction of these retrotransposons is mathematically represented in the form of classical "predator-prey" equations, where LINE-1 (considered as the "prey") produces the ORF2p protein, which contains a reverse transcriptase to insert new copies of retrotransposons into the genome, and SINE is considered as the "predator" since it uses LINE-1's ORF2p for its insertions. The model is a system of generalized Lotka-Volterra equations and reflects three types of competition: competition for ATP between retrotransposons and constitutive genes, competition for ribosomes between mRNAs of LINE-1 and constitutive genes, and competition for ORF2p between LINE-1 and SINE. Most parameter values in the model were estimated from the literature data [4]. The remaining part of the parameters were adjusted to obtain a stationary solution in which the values of the dynamic variables would correspond to the known estimates [4]. The model was implemented using Python.

Results: The resulting model has a stable stationary solution in which both LINE-1 and SINE retain low activity. It is assumed that this scenario is implemented in a cell under "normal conditions" and represents a reference solution. The existence of two other stationary solutions is shown, corresponding to the cases of SINE deactivation in the genome and complete deactivation of retrotransposons and their products in the cell. The regime in which SINE becomes inactive is characterized by a threefold decrease in the amount of ATP in the steady state due to increased energy consumption by the processes associated with LINE-1.

To find the parameters whose changes have the greatest impact on the ATP level in the cell, a parametric sensitivity analysis was performed. New values of the dynamic variables at the steady state were calculated when the most influential parameters associated with the retrotransposons were changed. It was shown that, among the retrotransposon-related parameters, the LINE-1 deactivation rate constant demonstrated the highest positive influence on the amount of available ATP molecules, and the LINE-1 maximum transcription rate exhibited the highest negative influence. Simultaneous perturbations in these parameters lead to a cumulative effect, so that a threefold change in these parameters relative to the reference values is sufficient for a threefold decrease in ATP levels. The same effect can be achieved by a fivefold change in each parameter separately. Under perturbation of these parameters, the leading source of ATP consumption changes from the translation in the reference state to the retrotransposon transcription.

We investigated how the initial concentration of retrotransposons in the cell influenced the instant level of free ATP molecules. A 100-fold increase in the initial amount of

LINE-1 leads to a threefold short-term decrease in the ATP level. For a 1000-fold increase in the initial amount of LINE-1, a more than threefold decrease in the ATP level persists for 13 minutes, and the maximum ATP level reduction becomes tenfold.
Conclusions: The developed mathematical model reflects the energy balance in the cell under conditions of competition for cellular resources under retrotransposon activity. We showed the possibility of a significant influence on cellular resources by transposons under modifications of their transcription rate and silencing mechanisms. The described effect can be used to initiate programs of cell death in cancer cells. The modeling results indicate the fundamental possibility of implementing scenarios for the destruction of tumor cells by increasing the load on the energy balance from activated transposons.
Funding: The research was financially supported by the Russian Science Foundation (project No. 23-24-00153).

Список литературы/References

1. Burns K. Transposable elements in cancer. *Nat Rev Cancer*. 2017;17:415-424. doi 10.1038/nrc.2017.35
2. Zhao Y., Oreskovic E., Zhang Q., Lu Q., Gilman A., Lin Y.S., He J., Zheng Z., Lu J.Y., Lee J., Ke Z., Ablaeva J., Sweet M.J., Horvath S., Zhang Z., Nevo E., Seluanov A., Gorbunova V. Transposon-triggered innate immune response confers cancer resistance to the blind mole rat. *Nat Immunol*. 2021;22(10):1219-1230. doi 10.1038/s41590-021-01027-8
3. Skulachev V.P. Bioenergetic aspects of apoptosis, necrosis and mitoptosis. *Apoptosis*. 2006;11(4):473-485. doi 10.1007/s10495-006-5881-9
4. Search BioNumbers – The Database of Useful Biological Numbers. Просмотрено: 14 май 2024 г. [Онлайн]. Доступно на: https://bionumbers.hms.harvard.edu/search.aspx

doi 10.18699/bgrs2024-2.1-09

Реконструкция генных сетей дифференциально экспрессирующихся генов в структурах мозга ручных и агрессивных крыс с использованием веб-сервиса ANDSystem

Яцык И.В.[1, 3]*, Чадаева И.В.[1], Ощепков Д.Ю.[1], Иванисенко В.А.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

[3] *Новосибирский государственный университет, Новосибирск, Россия*

* *woikin88@mail.ru*

Ключевые слова: агрессивное поведение; ANDSystem; ассоциативные генные сети; дифференциально экспрессирующиеся гены

Мотивация и цель: Агрессивное поведение животных изучается уже давно и занимает ведущее место в научных исследованиях основ поведения. Выявление генетических детерминант агрессивного поведения имеет большое значение, поскольку в человеческом обществе повышенная агрессивность – одна из основных проблем социальных отношений. Также важно то, что повышенная агрессивность наблюдается при многих психиатрических и нейродегенеративных заболеваниях, поэтому задача уменьшения проявлений агрессии весьма актуальна.

В проявлении агрессивного поведения велика роль наследственных факторов. Современные исследования говорят о том, что агрессивный тип поведения имеет полигенную детерминацию. В настоящей работе были продолжены исследования, направленные на изучение генетических детерминант агрессивного поведения. Целью работы является реконструкция и анализ генных сетей дифференциально экспрессирующихся генов и компьютерный анализ топологии полученных генных сетей с помощью веб-сервиса ANDSystem. Также были выявлены общие и тканеспецифичные биологические процессы по данным анализа генных сетей для анализируемых структур мозга.

Методы и алгоритмы: Работу проводили на данных секвенирования тканей мозга взрослых самцов серых крыс (*Rattus norvegicus*), которые были селекционированы в Институте цитологии и генетики СО РАН (Новосибирск). Селекция проводилась в течение 90 поколений на высокий уровень агрессии по отношению к человеку.

Для построения списка дифференциально экспрессирующихся генов (ДЭГ) у крыс с агрессивным поведением использованы данные РНК-секвенирования четырех отделов мозга (гипоталамус, гиппокамп, покрышка среднего мозга и серое вещество периакведуктума). Библиотеки чтений транскриптов от анализируемых крыс доступны на сайте NCBI (https://www.ncbi.nlm.nih.gov/sra) под индексом PRJNA668014.

Реконструкция генных сетей ДЭГ была произведена с использованием метода автоматического извлечения знаний из текстов научных публикаций (textmining) и баз данных о молекулярно-генетических взаимодействиях. Автоматическая реконструкция генных сетей проведена с помощью системы ANDSystem [1]. Расширение списка ДЭГ новыми генами, связанными с ними функционально, также осуществлено с помощью системы ANDSystem.

Чтобы построить генные сети, описывающие регуляторные пути, в которых участвуют белки, кодируемые дифференциально экспрессирующимися генами, был использован «Мастер путей», имеющийся в программе ANDVisio. С его помощью были составлены 5 шаблонов для построения регуляторных молекулярно-генетических путей 5 типов. С помощью этих шаблонов были описаны различные типы регуляторных путей молекулярно-генетических взаимодействий.
Схемы шаблонов указаны ниже:

1. Белки, кодируемые дифференциально экспрессирующимеся генами, – белок-белковые взаимодействия – все белки крысы.
2. Белки, кодируемые дифференциально экспрессирующимеся генами, – регуляция активности/стабильности/протеолиза/посттрансляционных модификаций/транспорта – все белки крысы.
3. Белки, кодируемые дифференциально экспрессирующимеся генами, – регуляция экспрессии генов – все гены крысы – экспрессия – все белки крысы.
4. Белки, кодируемые дифференциально экспрессирующимеся генами, – регуляция экспрессии генов – все гены крысы – экспрессия – все белки крысы – регуляция экспрессии генов – все гены крысы – экспрессия – все белки крысы.
5. Белки, кодируемые дифференциально экспрессирующимеся генами, – регуляция экспрессии генов – все гены крысы – экспрессия – все белки крысы – регуляция активности/стабильности/протеолиза/посттрансляционных модификаций/транспорта – все белки крысы.

После построения генных сетей по перечисленным шаблонам были рассмотрены их статистические параметры. Проанализированы топологические свойства полученных генных сетей по центральности, а также по посредничеству и степени вершин.
Результаты: С помощью шаблонов, указанных в разделе «Методы и алгоритмы», были построены регуляторные молекулярно-генетические пути, в которых участвуют белки, кодируемые дифференциально-экспрессирующимеся генами. Были найдены дополнительные гены, участвующие в молекулярно-генетических путях. В их числе оказались гены, ассоциированные с агрессией, например, BDNF, NPY, гены серотониновой системы и др. Из полученных молекулярно-генетических путей были построены как общая генная сеть, так и тканеспецифичные генные сети взаимодействия дифференциально-экспрессирующихся генов.
Был проведен статистический анализ реконструированных генных сетей. В ходе этого анализа были выявлены наиболее важные участники полученных сетей. Многие из них, такие как нейротрофический фактор BDNF, уже давно изучаются в исследованиях, связанных с агрессивным поведением. Роль других, таких как белок теплового шока HP74 или транскрипционный фактор P53, в проявлении агрессивного поведения изучена недостаточно хорошо.
Анализ полученных генных сетей с помощью веб-инструмента DAVID показал высокую представленность биологических процессов, связанных с работой нервной системы: нейроактивное лиганд-рецепторное взаимодействие, сигнальный путь цАМФ, кальциевый сигнальный путь и т. д.
Выводы: С помощью ANDSystem были реконструирована как общая генная сеть, так и тканеспецифичные генные сети взаимодействия дифференциально экспрессирующихся в четырех структурах мозга агрессивных и ручных серых крыс генов. Был проведен статистический анализ реконструированных генных сетей, с помощью которого выявлены наиболее важные участники полученных сетей. Анализ

представленности биологических процессов показал высокую представленность процессов, связанных с работой нервной системы.
Финансирование: Исследование поддержано бюджетным проектом № FWNR-2022-0020.

Reconstruction of gene networks of differentially expressed genes in brain structures of tame and aggressive rats using the ANDSystem web service

Yatsyk I.V.[1]*, Chadaeva I.V.[1, 2], Oshchepkov D.Y.[1, 2], Ivanisenko V.A.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *woikin88@mail.ru*

Key words: aggressive behavior; ANDSystem; associative gene networks; differentially expressed genes

Motivation and Aim: Aggressive behavior in animals has been studied for a long time and occupies a leading position in scientific research on the foundations of behavior. Identifying the genetic determinants of aggressive behavior is of great importance since increased aggressiveness is one of the major problems in social relations in human society. Additionally, heightened aggression is observed in many psychiatric and neurodegenerative diseases, making the task of reducing manifestations of aggression highly relevant.

Hereditary factors play a significant role in the manifestation of aggressive behavior. Modern studies suggest that the aggressive behavior type has a polygenic determination. My work continues the research aimed at studying the genetic determinants of aggressive behavior. The goal of this work is the reconstruction and analysis of gene networks of differentially expressed genes and the computer analysis of the topology of the resulting gene networks using the ANDSystem web service. Common and tissue-specific biological processes were also identified based on gene network analysis for the analyzed brain structures.

Methods and Algorithms: The work was carried out on sequencing data from the brain tissues of adult male gray rats (Rattus norvegicus) selected at the Institute of Cytology and Genetics SB RAS in Novosibirsk. Selection was conducted over 90 generations for a high level of aggression towards humans.

To compile the list of differentially expressed genes (DEGs) in aggressive rats, RNA sequencing data from four brain regions (hypothalamus, hippocampus, midbrain tegmentum, and periaqueductal gray matter) were used. Transcript libraries from the analyzed rats are available on the NCBI website (https://www.ncbi.nlm.nih.gov/sra) under the accession number PRJNA668014.

The reconstruction of gene networks of DEGs was performed using the method of automatic knowledge extraction from scientific publications (text mining) and molecular-genetic interaction databases. The automatic reconstruction of gene networks was carried out using the ANDSystem [1]. The expansion of the DEG list with new functionally related genes was also accomplished using the ANDSystem.

To construct gene networks describing regulatory pathways involving proteins encoded by differentially expressed genes, the "Pathway Master" available in the ANDVisio program was used. Five templates for constructing regulatory molecular-genetic pathways of five types were created using this tool. These templates described various types of regulatory pathways of molecular-genetic interactions.
The template schemes are listed below:

1. Proteins encoded by differentially expressed genes – protein-protein interactions – all rat proteins.
2. Proteins encoded by differentially expressed genes – regulation of activity/stability/proteolysis/post-translational modifications/transport – all rat proteins.
3. Proteins encoded by differentially expressed genes – gene expression regulation – all rat genes – expression – all rat proteins.
4. Proteins encoded by differentially expressed genes – gene expression regulation – all rat genes – expression – all rat proteins – gene expression regulation – all rat genes – expression – all rat proteins.
5. Proteins encoded by differentially expressed genes – gene expression regulation – all rat genes – expression – all rat proteins – regulation of activity/stability/proteolysis/post-translational modifications/transport – all rat proteins.

After constructing gene networks using the templates listed above, their statistical parameters were examined. The topological properties of the resulting gene networks were analyzed for centrality, betweenness, and degree of vertices.
Results: Using the templates specified in the "Methods and Algorithms" section, regulatory molecular-genetic pathways involving proteins encoded by differentially expressed genes were constructed. Additional genes involved in molecular-genetic pathways were identified. Among them were genes associated with aggression, such as BDNF, NPY, serotonin system genes, and others. From the obtained molecular-genetic pathways, both a general gene network and tissue-specific gene networks of interactions of differentially expressed genes were constructed.
A statistical analysis of the reconstructed gene networks was conducted. This analysis identified the most important participants in the resulting networks. Many of them, such as the neurotrophic factor BDNF, have long been studied in research related to aggressive behavior. The roles of others, such as the heat shock protein HP74 or the transcription factor P53, in the manifestation of aggressive behavior are not yet well understood. An analysis of the obtained gene networks using the DAVID web tool showed a high representation of biological processes related to the functioning of the nervous system: neuroactive ligand-receptor interaction, cAMP signaling pathway, calcium signaling pathway, etc.
Conclusion: Using ANDSystem, both a general gene network and tissue-specific gene networks of interactions of differentially expressed genes in 4 brain structures of aggressive and tame gray rats were reconstructed. A statistical analysis of the reconstructed gene networks was conducted, identifying the most important participants in the resulting networks. An analysis of the representation of biological processes showed a high representation of processes related to the functioning of the nervous system.
Funding: The study is supported by budgetary project No. FWNR-2022-0020.

Список литературы/References

1. Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. BMC Bioinformatics. 2019;20(Suppl. 1):34. doi 10.1186/s12859-018-2567-6

doi 10.18699/bgrs2024-2.1-10

Identification of co-expressed genes during macrophage polarization and reconstruction of their regulatory mechanisms using artificial intelligence methods

Antropova E.A.[1]*, Myakinkov I.[1, 2], Demenkov P.S.[1, 2], Ivanisenko V.A.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *nzhenia@bionet.nsc.ru*

Key words: immune system; macrophages; gene networks; expression regulation

Motivation and Aim: Macrophages play a key role in maintaining homeostasis of the body and its protection from damage and infections, being the central controlling element of immune responses [1, 2]. In response to incoming signals, macrophages are able to provide phenotypic adaptation from a pro-inflammatory phenotype (M1) to wound healing and tissue regeneration (group of M2 phenotypes), thus exhibiting plasticity [3]. In publications there are observations about the relationship between individual macrophage phenotypes and certain pathologies, or the relationship between the outcome of a disease and a certain macrophage phenotype. For example, studies have shown that patients with high-grade serous papillary ovarian cancer who have a higher M1/M2 macrophage ratio had significantly better overall survival [4]. Investigation of the molecular mechanisms underlying the functions of different macrophage phenotypes may allow the manipulation of immune responses. The goal of our work was to identify groups of co-expressed genes in different macrophage phenotypes and their subsequent analysis.
Methods and Algorithms: To analyze transcriptomic data, publicly available single cell RNA-seq data obtained on macrophages under different experimental conditions were taken. To detect co-expressed genes, a software package based on artificial intelligence methods was developed and applied.
Results: Single cell RNA-seq transcriptomic data was analyzed and differentially expressed genes were identified in macrophages under different conditions corresponding to the phenotype M1 (after adding LPS or IFN-γ+LPS), M2a (after adding IL-4+IL-13) and M2c (after adding IL-10), as well as M0 (macrophages without inducers), used as a control. Next, using the developed software package using artificial intelligence methods, groups of co-expressing genes characteristic of each phenotype were found. The resulting groups of coexpressed genes were analyzed and compared with genes accepted in the scientific literature as markers of each phenotype, similarities and differences were found. Biological processes and pathologies associated with the co-expressing groups of genes were found.
Conclusion: A new approach to the study of co-expressed genes based on artificial intelligence has revealed differences in the molecular mechanisms that provide different functions of different macrophage phenotypes.
Funding: The study is supported by the budget project FWNR-2022-0020.

References

1. Dash S.P., Gupta S., Sarangi P.P. Monocytes and macrophages: Origin, homing, differentiation, and functionality during inflammation. *Heliyon*. 2024;10(8):e29686. doi 10.1016/j.heliyon.2024.e29686
2. Mills C.D. M1 and M2 Macrophages: Oracles of Health and Disease. *Crit Rev Immunol*. 2012;32(6):463-488. doi 10.1615/critrevimmunol.v32.i6.10
3. Gurvich O.L., Puttonen K.A., Bailey A., Kailaanmäki A., Skirdenko V., Sivonen M., Pietikäinen S., Parker N.R., Ylä-Herttuala S., Kekarainen T. Transcriptomics uncovers substantial variability associated with alterations in manufacturing processes of macrophage cell therapy products. *Sci Rep*. 2020;10(1):14049. doi 10.1038/s41598-020-70967-2
4. Macciò A., Gramignano G., Cherchi M.C., Tanca L., Melis L., Madeddu C. Role of M1-polarized tumor-associated macrophages in the prognosis of advanced ovarian cancer patients. *Sci Rep*. 2020;10(1):6096. doi 10.1038/s41598-020-63276-1

doi 10.18699/bgrs2024-2.1-12

Challenges of metabolic pathways proteogenomic mapping

Chudinov I.K.[1, 2]*, Butenko I.O.[1], Vvedenskii A.V.[1], Korneenko E.V.[1, 3], Bondar S.V.[1], Speranskaya A.S.[1]

[1] *Scientific Research Institute for Systems Biology and Medicine, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology (National Research University), Moscow, Russia*

[3] *Saint Petersburg Pasteur Institute, Federal Service on Consumer Rights Protection and Human Well-Being Surveillance, St. Petersburg, Russia*

* *ikchudinov@gmail.com*

Key words: genomics; proteomics; proteogenomics; metabolic pathways

Motivation and Aim: Dried multispecies biotechnological bacterial preparations, known as starter cultures, are used in the production of fermented dairy products. According to the European Food Safety Authority (EFSA) approach, each bacterial species has additional specifications for safe use based on existing knowledge. Therefore, appropriate methods of investigation are required for bacterial typing and detection of potential hazards. Well-established bacterial typing approaches, including 23S RNA sequencing or multilocus typing, may be insufficient for resolution of different close strains and lack insight into metabolic capabilities of these preparations. In this work, we tested the applicability of a proteogenomic approach for metabolic pathway mapping in *Lactococcus lactis*.

Methods and Algorithms: Bacterial isolates were obtained from lactic acid starter cultures of six different manufacturers. Selected isolates were grown on Petri dishes and collected for DNA isolation and bulk growth.

DNA sequencing was performed using PE100 on the MGI platform according to the manufacturer's recommendations. Genome assembly and annotation was performed using pipeline bacass [1].

Five biological replicates for each of six isolates were collected for proteomic analysis. Samples were lysed with sonication, total protein was precipitated, resolubilized, disulfide bonds were reduced, free tiol groups were alkilated and total protein fraction was digested with trypsin, dried and resolubilized. Each peptide sample was subjected to panoramic label-free data-dependent LC-MS analysis.

MSFragger [2] with Percolator [3] and Sage [4] with Mokapot [5] were used independently for spectral data evaluation with sequencing-based annotations and the reference proteome UP000326269 [6]. Pathway reconstruction was performed using QuickGO [7]. Subsequent data processing was performed with in-house Python scripts.

Results: Six draft genomes were obtained and annotated from the sequencing data. All isolates were typed as *Lactococcus lactis* subsp. *lactis bv. diacetylactis* strain SD96. Each of the samples is characterized with a shotgun proteomic data in multiple biological replicates. Peptides from more than 1000 proteins have been detected within all spectral identification approaches of the samples. Among the detected proteins, most of them are related to translation, peptidoglycan biosynthesis process and regulation of DNA-templated transcription processes. It was not possible to conclusively distinguish samples by characteristic protein signatures or metabolic pathways, but it is still possible to infer possible pathways in the culture of interest.

Conclusion: Existing approaches allow to determine the presence of genes and proteins that characterize the main metabolic pathways, but are not always sufficient to separate bacterial strains or to infer the possibility of producing potentially dangerous metabolites. To the best of our knowledge, there are no "out-of-the-box" solutions that address the ongoing challenges of computational omics research.

Funding: The study is supported by the state project "Development of algorithms to identify novel, unique DNA or RNA sequences in metagenomes and their phenotypic characterization in vitro" (No. 12203090069-4).

References

1. Peltzer A., Straub D. et al. nf-core/bacass: v2.1.0 nf-core/bacass: "Navy Steel Swordfish" (2.1.0). *Zenodo*. 2023. doi 10.5281/zenodo.10036158
2. Kong A., Leprevost F., Avtonomov D. et al. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nat Methods*. 2017;14:513-520. doi 10.1038/nmeth.4256
3. The M., MacCoss M.J., Noble W.S., Käll L. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J Am Soc Mass Spectrom*. 2016;27(11):1719-1727. doi 10.1007/s13361-016-1460-7
4. Lazear M.R. Sage: An Open-Source Tool for Fast Proteomics Searching and Quantification at Scale. *J Proteome Res*. 2023;22(11):3652-3659. doi 10.1021/acs.jproteome.3c00486
5. Fondrie W.E., Noble W.S. mokapot: Fast and Flexible Semisupervised Learning for Peptide Detection. *J Proteome Res*. 2021;20(4):1966-1971. doi 10.1021/acs.jproteome.0c01010
6. Dorau R., Chen J., Jensen P.R., Solem C. Complete Genome Sequence of *Lactococcus lactis* subsp. *lactis* bv. diacetylactis SD96. *Microbiol Resour Announc*. 2020;9(3):e01140-19. doi 10.1128/MRA.01140-19.
7. Binns D., Dimmer E., Huntley R., Barrell D., O'Donovan C., Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*. 2009;25(22):3045-3046. doi 10.1093/bioinformatics/btp536

doi 10.18699/bgrs2024-2.1-13

In silico analysis of the system determining the morphogenesis of *Drosophila* mechanoreceptors

Furman D.P.[1, 2]*, Bukharina T.A.[1, 2]**

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

** furman@bionet.nsc.ru, ** bukharina@bionet.nsc.ru*

Key words: central regulatory circuit; gene network; mathematical model; computer modeling; *Drosophila*; achaete-scute complex; mutations

Motivation and Aim: Identification of the mechanisms underlying the genetic control of spatial structure formation is among the relevant tasks of developmental biology. Both experimental and theoretical approaches and methods are used for this purpose, including gene network methodology [1], as well as mathematical and computer modeling. Reconstruction and analysis of the gene networks that provide the formation of traits allow us to integrate the existing experimental data and to identify the key links and intra-network connections that ensure the function of networks. Mathematical and computer modeling is used to obtain the dynamic characteristics of the studied systems and to predict their state and behavior. The numerical experiments conducted with the help of mathematical models allow the potential operation modes of a system to be examined, its future states to be forecasted, and its new functions to be predicted by changing parameters or adding new assumptions. In many cases, modeling is the only way to clarify the processes taking place in a system when their characteristics cannot be directly measured in a biological experiment.

An example of the spatial morphological structure is the *Drosophila* bristle pattern with a strictly defined arrangement of its components, mechanoreceptors. The mechanoreceptor develops from a single sensory organ parental cell (SOPC), which is isolated from the ectoderm cells of the imaginal disk. It is distinguished from its surroundings by the highest content of proneural proteins (ASC), the products of the *achaete-scute* proneural gene complex (*AS-C*). The SOPC status is determined by the gene network we earlier reconstructed and the *AS-C* is the key component of this network. *AS-C* activity is controlled by its subnetwork – the central regulatory circuit (CRC), comprising seven genes: *AS-C*, *hairy*, *senseless* (*sens*), *charlatan* (*chn*), *scratch* (*scrt*), *phyllopod* (*phyl*), and *extramacrochaete* (*emc*), as well as the corresponding proteins. In addition, the CRC includes the accessory proteins Daughterless (DA), Groucho (GRO), Ubiquitin (UB), and Seven-in-absentia (SINA) [2].

Although the morphogenesis of mechanoreceptors has been long studied, it is still far from an exhaustive description. It is only qualitatively characterized: the players in this process (genes and proteins) are known and the general concept of their interaction is formed; however, most of the quantitative parameters as well as a relative contribution of the involved genes have not been experimentally determined. The goal of this work was to construct a mathematical model of CRC operation taking into account the roles of the constituent genes that would comprehensively describe the intracellular events in

presumptive SOPC determining the dynamics of ASC content and to perform the computer experiments for verifying the model stability and its compliance with experimental data.

Methods and Algorithms: The proposed dynamical model of *AS-C* activity is described with a system of seven ordinary differential equations (ODE) [3]. The variables in this system are the concentrations of CRC proteins in the cell: $x(t)$ is the content of ASC; $y(t)$, of Hairy; $E(t)$, of Extramacrochaete; and $z(t)$, $u(t)$, $w(t)$, and $p(t)$, the concentrations of Senseless, Scratch, Charlatan, and Phyllopod, respectively. To take into account the mutations of the genes that compose the CRC, the model contains non-negative coefficients k_x, k_y, k_e, and so on reflecting the degrees of the effect of mutations on the synthesis of the corresponding proteins. The values of these coefficients do not exceed unity; $k = 1$ corresponds to the normal operation of a gene; and $k = 0$ denotes a complete inactivation of a gene and the absence of the corresponding protein.

The specially developed web application (https://gene-nets-simulation.shinyapps.io/crc-asc-modeler/) allows the CRC operation modes to be simulated for different values of the ODE system parameters and the results of these numerical experiments to be visualized as plots [4].

Results: The proposed model allowed us to advance from a purely qualitative description of the system controlling the content of ASC proteins and to clarify some of its quantitative characteristics unknown earlier.

In particular, our numerical experiments suggest that the cell is determined as an SOPC when the ASC content increases approximately 2.5-fold relative to the initial level in the cells of proneural cluster. Individual elements of the circuit have different effects on the content of ASC proteins in the presumptive cell of mechanoreceptor. *AS-C*, the key CRC component, and the mutations that decrease the ASC content by more than 40 % have the most significant effect and cause the prohibition of SOPC segregation. As for the mutations in the remaining genes of the circuit, they change the level of ASC proteins to different degrees, with the most pronounced effects of mutations in *emc* and *hairy* genes. Thus, the model demonstrates that the CRC as a system is sensitive to the changes in internal interactions and its robust operation, providing a certain dynamics of the level of ASC proteins, requires a concerted work of all components constituting the regulatory circuit.

Funding: The authors are sincerely grateful to V.P. Golubyatnikov and A.A. Akinshin for their assistance in creating the model, helpful advice, and criticism. The study is supported by the budget project No. FWNR-2022-0020.

References

1. Kolchanov N.A., Ignat'eva E.V., Podkolodnaya Q.A., Likhoshvai V.A., Matushkin Y.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2013;17(4/2):833-849 (in Russian)
2. Furman D.P., Bukharina T.A. Genetic regulation of morphogenesis of *Drosophila melanogaster* mechanoreceptors. *Russ J Dev Biol*. 2022;53(4):239-251. doi 10.1134/S1062360422040038
3. Furman D.P., Bukharina T.A., Golubyatnikov V.P. Central Regulatory Circuit of the Drosophila mechanoreceptor morphogenesis system: effects of mutations. *J Appl Ind Math*. 2023;17(3):535-543. doi 10.1134/S1990478923030079
4. Bukharina T.A., Akinshin A.A., Golubyatnikov V.P., Furman D.P. Mathematical and Numerical Models of the Central Regulatory Circuit of the Morphogenesis System of Drosophila. *J Appl Ind Math*. 2020;14(2):249-255. doi 10.1134/S1990478920020040

doi 10.18699/bgrs2024-2.1-14

Analysis of the evolution of the human genes encoding cell surface receptors involved in the regulation of appetite using phylostratigraphic age and divergence indexes

Ignatieva E.V.*, Ivanov R.A., Lashin S.A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** eignat@bionet.nsc.ru*

Key words: regulation of appetite; cell surface receptors; hunger; evolution; phylostratigraphic analysis; gene age; gene variability

Motivation and Aim: Appetite is a drive to consume food. It is one of the most primitive instincts promoting survival and it exists in all higher life-forms. Appetite serves to regulate adequate energy intake to maintain metabolic needs, although it may be stimulated by appealing foods even when hunger is absent. Dysregulation of appetite can lead to human diseases (bulimia nervosa and overeating (both leading to obesity), anorexia nervosa, cachexia, etc.) [1]. Understanding the evolutionary mechanisms of human diseases, especially those related to lifestyle changes that have occurred over the past 100–200 years, is of fundamental and applied importance. It is also very important to identify relationships between the evolutionary characteristics of genes in gene networks and the resistance of these networks to changes caused by mutations. We have previously analyzed the evolutionary characteristics of genes involved in networks from the KEGG Pathway, Human Diseases database describing various human diseases [2]. The features of these networks have been identified both in the context of the phylostratigraphic age (time of gene origin) of their constituent genes and the nature of the selection to which they are subjected (stabilizing or driving) [2]. On the other hand, we have previously performed a functional analysis of genes involved in the regulation of appetite and body weight [3, 4]. When analyzing a set of 105 genes involved in appetite regulation, a statistically significant over-enrichment of genes specifically expressed in the brain was found. It was also revealed that a substantial proportion of genes (~45 %) in this set were genes encoding cell surface receptors. Many of these receptors belonged to the superfamily of G-protein-coupled receptors (GPCRs). Thus, the aim of the current study is to identify the distinctive features of human genes encoding cell surface receptors involved in appetite regulation using the phylostratigraphic age index (PAI) and divergence index (DI).
Methods and Algorithms: A set of human receptor genes whose orthologues were involved in appetite regulation in model organisms was formed based on the analysis of scientific publications. Evolutionary characteristics of genes (phylostratigraphic age index (PAI) and divergence index (DI)) have been calculated as it was described in [2] taking into account sequences of orthologous genes that are 50 % or more identical to one under consideration. To identify genes involved in the control of biological processes related to development we used annotation of genes by GO terms obtained from DAVID tool [5].

Results: We found 80 human genes orthologous to genes identified in model organisms and encoding cell surface receptors. It turned out that the set of genes under consideration contains an increased number of genes (37) with the same phylostratigraphic age (PAI=6, the stage of vertebrate divergence), and almost all of these genes (35 out of 37) belong to the superfamily of G-protein coupled receptors. Apparently, the synchronized evolution of such a large group of genes (37 genes out of 80) is associated with the development of the brain as a separate organ in the first vertebrates.
When studying the distribution of genes from the same set by DI values, a significant enrichment with genes that had low DIs was revealed. Eleven genes (*GPR26*, *NPY1R*, *GHSR*, *CNR1*, *ADIPOR1*, *DRD1*, *MCHR1*, *ADCYAP1R1*, *NPY2R*, *GPR171*, and *NPBWR1*) had an extremely low DIs (less than 0.05). Such low values of DI indicate that these genes are very likely exposed to stabilizing selection. The lowest DI was found for *GPR26*. Due to the fact that the endogenous ligands for the GPR26 receptor had not yet been identified yet, this gene seems to be an extremely interesting object for further theoretical and experimental research. It was also found that the group of genes with low DIs was enriched with genes that were related to development.
Conclusion: We believe that the features of genes encoding cell surface receptors we have identified using PAI and DI evolutionary metrics can be a the starting point for further evolutionary analysis of the gene network regulating appetite.
Funding: The study is supported by publicly funded project No. FWNR-2022-0020 of the Federal Research Center ICG SB RAS.

References

1. Gibson D., Workman C., Mehler P.S. Medical Complications of Anorexia Nervosa and Bulimia Nervosa. *Psychiatr Clin North Am.* 2019;42(2):263-274. doi 10.1016/j.psc.2019.01.009
2. Mustafin Z.S., Lashin S.A., Matushkin Yu.G. Phylostratigraphic analysis of gene networks of human diseases. *Vavilov Journal of Genetics and Breeding*. 2021;25(1):46-56. doi 10.18699/VJ21.006
3. Ignatieva E.V., Afonnikov D.A., Rogaev E.I., Kolchanov N.A. Human genes controlling feeding behavior or body mass and their functional and genomic characteristics: a review. *Vavilov Journal of Genetics and Breeding*. 2014;18(4/2):867-875 (in Russian)
4. Ignatieva E.V., Afonnikov D.A., Saik O.V., Rogaev E.I., Kolchanov N.A. A compendium of human genes regulating feeding behavior and body weight, its functional characterization and identification of GWAS genes involved in brain-specific PPI network. *BMC Genet*. 2016;17(Suppl. 3):158. doi 10.1186/s12863-016-0466-2.
5. Sherman B.T., Hao M., Qiu J., Jiao X., Baseler M.W., Lane H.C., Imamichi T., Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* 2022;50(W1):W216-W221. doi 10.1093/nar/gkac194

doi 10.18699/bgrs2024-2.1-15

Reconstruction and computer analysis of the structural and functional organization of the gene network regulating cholesterol biosynthesis in humans and the evolutionary characteristics of genes involved in the network

Ignatieva E.V.[1]*, Mikhailova A.D.[2], Lashin S.A.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *eignat@bionet.nsc.ru*

Key words: cholesterol biosynthesis; gene network; feedback loops; evolution; gene variability; phylostratigraphic analysis; gene age; stabilizing selection; SREBP

Motivation and Aim: Cholesterol is a structural component of cell membranes, a precursor to steroid hormones and other biochemical substances in the animal body. Cholesterol is consumed by animals with food, and is also synthesized in the cells *de novo*. Previously, the gene network of regulating intracellular cholesterol level was reconstructed [1]. This network comprised a number of regulatory loops, each of which included transcription factors of the SREBP family (SREBP-1 and SREBP2). The activity of transcription factors of the SREBP family (SREBPs) is regulated inversely depending on the intracellular cholesterol level. The functioning of this mechanism is carried out by several proteins, which, in combination with SREBPs, may be called "cholesterol sensor" (that is, SCAP, INSIG1, INSIG2, SREBP1, SREBP2). The other mechanisms regulating SREBPs functioning are known, due to which a precise modulation of SREBPs activity is carried out depending on the state of the cell and the organism as a whole [1]. It is also known that elevated cholesterol level may be a risk factor for the development of cardiovascular diseases (atherosclerosis, coronary heart disease), and may increase the severity of other pathological processes. Systematization of data on the molecular mechanisms controlling the activity of SREBPs and cholesterol biosynthesis in the format of a gene network and obtaining new knowledge about the gene network as a single object is extremely important in the context of understanding the involvement of this system in the diseases progression. The aim of the current study is to build an extended gene network regulating cholesterol biosynthesis in humans and to analyze the structural and functional organization of the network and the evolutionary characteristics of genes.

Methods and Algorithms: We used ANDSystem [2] to reconstruct and to analyze the gene network regulating cholesterol biosynthesis. Evolutionary metrics of genes (phylostratigraphic age index (PAI) and divergence index (DI)) have been calculated as it was described in [3] taking into account sequences of orthologous genes that are 50 % or more identical to one under consideration.

Results: Using ANDSystem we built a new extended version of the gene network regulating cholesterol biosynthesis. It includes (1) proteins that make up the "cholesterol sensor" and the genes encoding them; (2) proteins regulating the activity of genes and pro-

teins of the “cholesterol sensor”; (3) cholesterol and enzymes involved in its biosynthesis; (4) genes regulated by transcription factors of the SREBP family, and the proteins encoded by them.

Regulatory circuits have been identified that control the activity of SREBPs. All these regulatory circuits include proteins encoded by the SREBPs target genes. The four shortest circuits show the regulatory effects of PPARG or NR0B2/SHP1 or FAS or LPIN1 on the activity of SREBP-1 or SREBP-2. Within the other five circuits, the regulation of SREBP-1 or SREBP-2 activity is carried out indirectly, due to the effects of FAS or PPARG or AR on the proteins functioning within “cholesterol sensor” (INSIG1 or SCAP), which, in turn, affect the activity of SREBP-1 or SREBP-2.

The analysis of the evolutionary characteristics of genes using PAI and DI metrics revealed that: (1) the genes involved in the gene network regulating cholesterol biosynthesis are characterized by decreased PAI values compared to the set of all human protein-coding genes, that may mean that they are on average more “ancient”, in addition, these genes have significantly decreased DI values (for most of them DI is less than 0.4), indicating a stronger pressure of stabilizing selection on these genes compared to the pressure of stabilizing selection on all human protein-coding genes as a whole; (2) the subgroup of genes, encoding enzymes of cholesterol biosynthesis contains significantly more genes with a greater phylostratigraphic age (PAI $\leq$ 2, cellular organisms (the root of the phylogenetic tree), and the stages of eukaryota and metazoa divergence) than protein-coding genes as a whole, that means that they are more “ancient”; (3) the subgroup of genes, encoding regulatory proteins that affect cholesterol sensor activity, contains significantly more genes with a greater phylostratigraphic age (PAI < 5, all stages before vertebrata divergence) than protein-coding genes as a whole, that means that they are more “ancient”; (4) the subgroup of genes, encoding regulatory proteins that affect cholesterol sensor, contains significantly more genes having low DI values (less than 0.4), indicating a stronger pressure of stabilizing selection on these genes compared to the pressure of stabilizing selection on all human protein-coding genes as a whole.

Conclusion: Analysis of the gene network regulating cholesterol biosynthesis has revealed a complex nature of the molecular-genetic mechanisms controlling this process. We identified new genes involved in feedback loops and revealed features of evolutionary characteristics of individual subgroups of genes involved in the network. Taking into account these findings may be useful in the search for new pharmacological targets for the treatment of pathologies associated with cholesterol level deviations.

Funding: The study is supported by publicly funded project No. FWNR-2022-0020 of the Federal Research Center ICG SB RAS.

References

1. Kolchanov N.A., Ignatieva E.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin Yu.G. Gene Networks *Vavilov Journal of Genetics and Breeding*. 2013;17(4/2):833-850
2. Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl. 1):34. doi 10.1186/s12859-018-2567-6
3. Mustafin Z.S., Lashin S.A., Matushkin Yu.G. Phylostratigraphic analysis of gene networks of human diseases. *Vavilov Journal of Genetics and Breeding*. 2021;25(1):46-56. DOI 10.18699/VJ21.006

doi 10.18699/bgrs2024-2.1-16

Automated reconstruction and analysis of gene networks using cognitive system ANDSystem

Ivanisenko V.A.[1, 2, 3]*, Demenkov P.S.[1, 2, 3], Ivanisenko T.V.[1, 2, 3], Antropova E.A.[1], Makarova A.L.A.[1, 2, 3], Adamovskaya A.V.[1, 2, 3], Venzel A.S.[1, 2, 3], Kolchanov N.A.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *salix@bionet.nsc.ru*

Cognitive computer systems are gaining increasing interest in the field of biomedicine for the reconstruction and analysis of gene networks. These systems offer a complete cycle of knowledge engineering, including entity recognition, relationship establishment, knowledge integration, and representation in the form of semantic networks or knowledge graphs. Additionally, they provide a range of tools for data analysis and interpretation, such as gene prioritization and identification of overrepresented biological processes or diseases.

We have developed a cognitive software and information system, ANDSystem [1], which considers 12 types of objects, including organisms, cells/tissues, diseases, genes, metabolites, drugs, etc. The system extracts knowledge about physical interactions, catalytic reactions, regulatory relationships, and more from scientific texts. The ANDSystem knowledge base, created based on large-scale automatic analysis of over 30 million PubMed abstracts and full-text articles from PMC, as well as hundreds of factual databases in the field of biology and biomedicine, is a unique resource containing formalized knowledge about more than 50 million interactions between molecular-genetic objects of different organisms. The developed system surpasses similar known systems in the world in terms of the number of considered types of relationships and types of objects.

ANDSystem has been used to solve a wide range of problems in the reconstruction of molecular mechanisms of diseases, the interpretation of omics data, and other problems in the field of biomedicine. In particular, we used ANDSystem to study the molecular genetic mechanisms of pathogen-host interaction.

Reconstruction of signaling pathways for the regulation of cellular biological processes by viral proteins can help in the search for new pharmacological targets. Using ANDSystem, we reconstructed associative gene networks describing the potential regulation of the apoptosis process by HCV viral proteins. Expanding the given gene network with consideration of information about aberrantly methylated genes in hepatocellular carcinoma enabled the prioritization of new promising pharmacological targets for liver cancer therapy. Furthermore, we reconstructed potential molecular mechanisms of dysfunction of signaling pathways in hepatocellular carcinoma.

The application of ANDSystem to the interpretation of metabolomics data, in particular, allowed us to investigate the regulatory role of SARS-CoV-2 non-structural proteins in the disruption of metabolic processes in patients diagnosed with COVID-19. The application of ANDSystem to the analysis of blood plasma metabolomic screening data from patients with postoperative delirium allowed us to reconstruct the molecular genetic mechanisms of sphingolipid metabolism disorders.

In conclusion, ANDSystem is a powerful cognitive system for the reconstruction and analysis of gene networks with a unique knowledge base that surpasses similar known systems in the world. The system has been successfully applied to solve a wide range of problems in the field of biomedicine.

Funding: The bioinformatics analysis was supported by the budget project, FWNR-2022-0020. The authors express their gratitude to Evgeniy V. Gaisler, Artem D. Rogachev, Nikita V. Basov, Andrey G. Pokrovsky for obtaining and analyzing the metabolomics data.

References

1. Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl. 1):34

doi 10.18699/bgrs2024-2.1-17

Analysis of context-specific genome-scale models for metabolism of fast- and slow-growing chicken breeds

Kaplan V.S.[1]*, Kulyashov M.A.[1], Volyanskaya A.R.[2], Yevshin I.S.[1], Shagimardanova E.I.[3], Gusev O.A.[3, 4, 5], Kolpakov F.A.[1], Akberdin I.R.[1]

[1] *Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[4] *Life Improvement by Future Technologies (LIFT) Center, Moscow, Russia*

[5] *Intractable Disease Research Center, Juntendo University, Tokyo, Japan*

* *kaplanrimi@gmail.com*

Key words: chicken metabolism; genome-scale metabolic models; context-specific models; differential expression genes; differential expression fluxes; flux balance analysis

Motivation and Aim: Synthetic biology, a rapidly advancing field of modern biology, focuses on constructing new organisms with specified properties through genome synthesis or deep reorganization using genetic engineering, bioengineering, systems biology and bioinformatics.Systems biology, integrating omics data, detailed analysis of gene interactions and metabolic modeling, offers a promising strategy for enhancing biotechnological properties [1]. Genome-scale metabolic models (GEMs) have become a powerful systems biology tool for understanding the metabolism of organisms. Furthermore, context-specific genome-scale models enable the study of metabolism under specific environmental or physiological conditions. By integrating transcriptomic data into a metabolic model, we can pinpoint specific metabolic reactions and genes associated with growth performance providing valuable insights for targeted breeding and nutritional interventions. Such a model makes it possible to predict *in silico* genetic modifications (gene knockouts, promoter modifications) required to increase the growth rate of a cell culture or even whole organism, the yield of the target product when the cell culture grows under optimal conditions and depending on different substrates or nutrition if it is an animal organism. Despite significant advancements in genetic selection and feeding practices, the molecular mechanisms underlying growth differences among chicken breeds remain poorly understood [2].

The primary aim of this study is to investigate the metabolic variability between fast- and slow-growing chicken breeds through the analysis of context-specific genome-scale models (csGEMs). We assume that the differential growth rates observed between fast- and slow-growing chicken breeds are governed by distinct metabolic signatures, characterized by variations in metabolic pathways, enzyme activities, and regulatory mechanisms. Essentially, we propose csGEMs that provide an opportunity to uncover unique metabolic features associated with growth performance in chickens.

Methods and Algorithms: The F2 generation from crosses between Russian White and Cornish breeds has a high diversity in growth rate, reaching a live weight of 1.5 to 2.5 kg by the fourth week. 12 slow-growing and 12 fast-growing chickens from the F2 generation were selected. We examined the transcriptome profiling of these 24 chickens by

CAGE-seq in heart, liver, kidney, brain, breast and leg muscles. CAGE-seq allows one the identification of transcriptional start sites and their activity level. 4.5 billion paired-end reads of 150 nucleotides in length were sequenced by MGI sequencer. Then we performed quality control on the raw sequencing data to filter out low-quality reads and adapters. A chicken pan-genome representing a GRCg6a reference assembly and an additive of 159 million base pairs as short contigs was created, from which we analyzed CAGE-seq data at the next steps. The reads were mapped to the chicken pangenome. The bowtie2 [3] was used for mapping RNA-seq data to the genome. We then normalized transcript abundance using the Transcript Per Million kilobase (TPM) method. These expression profiles were integrated into the iES1300 metabolic model [4] using the RIPTiDe algorithm [5], resulting in the generation of four context-specific models tailored to the respective chicken breeds and tissue types. We employed Escher [6] to construct metabolic maps for the context-specific models. The flux distribution ratios of these context-specific models, comparing fast to slow-growing breeds for each tissue, were simulated. Key and significantly different values of the same flux between two types of breeds were identified as differential expressed fluxes (DEFs). DEFs are characterized by flux changes between fast and slow breeds exceeding 1.5 times, disappearance, appearance, or reverse direction. Five types of DEFs were distinguished: Upper Regulated (accelerated), Down Regulated (decreased), Off (disappeared), On (appeared), and Reversed. Furthermore, we performed a comparative analysis of metabolic maps for context-specific models built on two different genome assemblies by looking for the ratio between the difference in flux values of the fast-growing to slow-growing pangenomic GRCg6a-based model to the same difference for the GRCg6a-based model. We also searched for different expression genes (DEGs) in fast and slow growing chickens based on CAGE-seq data and filtered them by p-adj<0.05. For this purpose, the number of reads in the identified transcriptional starts was counted and differential transcriptional starts were found by DESeq2 method [7], to which known genes were then compared according to Refseq annotation.

We use a refined GEM of a chicken cell named iES1300 [4], consisting of 2427 reactions, 2569 metabolites, and 1300 genes. We modified the model by adding important missing pseudo-reactions, metabolite formulas in order to test mass and charge balancing. Such manipulations of the model took place using Python library CobraPy [8]. Then we integrated TPM in the model using a specific tool, RIPTiDe [5] in order to construct context-specific models. The flux distributions of these models were compared between each other and analyzed; these fluxes were compared to previously obtained DEGs.

Results: We enhanced the model by incorporating 513 demand reactions, formulating metabolite formulas for 2569 metabolites, creating reactions, and ensuring mass-charge balance for 866 reactions from 916 non-balanced reactions. Evaluation through Memote test-suite [9] revealed an increase in the consistency score from 18 % in the original iES1300 model to 36 % in the modified version. Mass balance of the modified model became equal to 99.0 %, charge balance 98.8 % and unbounded flux in default medium 74.0 %, whereas mass balance of original model was 0.0 %, charge balance 97.2 % and unbounded flux in default medium 69.8 %. No changes were found in growth rate between the original and modified models, but secretion and uptake rates had pronounced differences. The constructed csGEMs were tested for correlation with transcriptome data and changes in growth rates compared to the GEM (Table 1).

Table 1. Correlation of csGEMs with TPM and the difference between the growth rate of the csGEMs and GEM

	Changes in growth rate in comparison with the GEM	r, correlation with transcriptomic data	p-value
fast-leg csGEM	No change	0.173	0.002
slow-leg csGEM	to ~0.0111 from ~0.0342	0.229	<0.001
fast-liver csGEM	No change	0.106	0.05
slow-liver csGEM	to ~0.0127 from ~0.0342	0.208	0.001

Visual and quantitative analysis of the metabolic map with simulation results enabled us to identify the strongest changes in the distribution of fluxes which were observed in the glycolysis and Krebs cycle (Fig. 1).

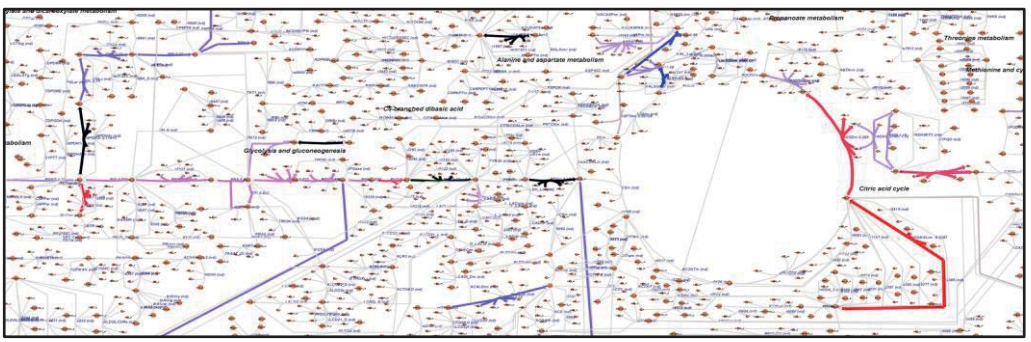


Fig. 1. A visualized part of the metabolic map with flux ratios for context-specific models based on GRCg6a and pangenomic GRCg6a assembled models

To check the quality of the FBA results we compared the results of DEFs and DEGs for leg muscle tissues. Additionally, the resulting DEFs for the GRCg6a- and pangenomic GRCg6a-based assemblies for leg muscle tissues were compared to each other. As a result, we found 28 out of 320 DEFs common in both types of context-specific models. Moreover, the resulting filtered DEGs were also compared to each other for GRCg6a- and pangenomic GRCg6a-based assemblies. It was identified from the analysis that 65 out of 131 common genes in both datasets are in the direction of DEGs for leg muscle tissues. We then compared DEGs with genes associated with DEF responses and found 67 of 250 genes are common in both types of the csGEMs. We also compared the DEFs between the original and modified models, and found 100 of 207 DEFs common in type. *Conclusion*: According to the Memote reports, we got a more qualitative model and made it more suitable for further biological interpretation. As we hypothesized, context-specific models of fast-growing muscle tissues significantly differ from model predictions for slow-growing tissues in terms of the growth rate as an objective function of flux balance analysis. Our pan-genomics assembly allowed us to refine the results of csGEMs, especially for crucial metabolic pathways such as glycolysis and the Krebs cycle. Our investigation delineated and clarified the key metabolic pathways and regulatory nodes contributing to the disparate growth rates observed in fast- and slow-growing chicken breeds.

All the results obtained have been make publicly available on gitlab: https://gitlab.sirius-web.org/collaboration/Chicken/gem_g.gallus2024

Funding: This study was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2021-1344).

References

1. Esvelt K.M., Wang H.H. Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol*. 2013;9:641. doi 10.1038/msb.2012.66
2. Volyanskaya A.R., Akberdin I.R., Kulyashov M.A. et al. A bird's-eye overview of molecular mechanisms regulating feed intake in chickens-with mammalian comparisons. *Anim Nutr*. 2024;17:61-74. doi 10.1016/j.aninu.2024.01.008
3. Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-9. doi 10.1038/nmeth.1923
4. Salehabadi E., Motamedian E., Shojaosadati S.A. Reconstruction of a generic genome-scale metabolic network for chicken: Investigating network connectivity and finding potential biomarkers. PLoS One. 2022;17(3):e0254270. doi 10.1371/journal.pone.0254270
5. Jenior M.L., Moutinho T.J. Jr., Dougherty B.V., Papin J.A. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Comput Biol*. 2020;16(4):e1007099. doi 10.1371/journal.pcbi.1007099
6. King Z.A., Dräger A., Ebrahim A., Sonnenschein N., Lewis N.E., Palsson B.O. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Comput Biol*. 2015;11(8):e1004321. doi 10.1371/journal.pcbi.1004321
7. Love M.I., Huber W., Anders S. Differential analysis of count data – the DESeq2 package. *Genome Biol*. 2014;15:10-1186
8. Ebrahim A., Lerman J.A., Palsson B.O., Hyduke D.R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol*. 2013;7:74. doi 10.1186/1752-0509-7-74
9. Lieven C., Beber M.E., Olivier B.G. et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol*. 2020;38(3):272-276. doi 10.1038/s41587-020-0446-y

doi 10.18699/bgrs2024-2.1-18

Systems Biology of next epidemics: 3 dimensions of emergence and ORCHESTRA experience

Kolodkin A.[1, 2]

[1] *Luxembourg Institute of Health, Strassen, Luxembourg*

[2] *University of Luxembourg, Esch-sur-Alzette, Luxembourg*

* *alexeykolodkin@gmail.com*

Key words: Systems Biology; Coplex Adaptive Systems; COPASI; CellDesigner, FAIR data management, COVID-19

Motivation and Aim: Systems Biology aims to understand biological emergence from the interactions of biomolecules, e. g. by integrating the knowledge about these interactions into a computer model and thereby reconstructing biological behavior in silico. In relation to the human, such an in silico replica of the whole body is the so-called Silicon Human. We can add medical aspects to this model and, using the patient's genome, transcriptome, and proteome data, parameterize a Silicon Human for any patient individually (Silicon Patient), for modeling any systems biological disease [1, 2], including so-called Long Covid caused by SARS-CoV-2 virus. On this level, personalized physiological behavior emerges from interactions between biomolecules, like in many other systems' biological models [3–5].

However, when talking about epidemics, there is also a second dimension of emergence. Interactions between susceptible, recovered, immune, or infected individuals lead to the emergence of epidemics. The chance of an individual being infected is now state-dependent on the spread of the virus in the population [6].

Moreover, there is also a third, level of emergence, where interactions between various players (researcher, clinical data manager, medical doctors, nurses, etc) collecting and processing different pieces of data lead to the emergence of the project. We should notice that the scientific concept is evolving during the study and thus the project itself is a Complex Adaptive System. The data collection process becomes state-dependent and data becomes live in a sense that they change with the evolution of the whole system. So, data should become not only Findable, Accessible, Interoperable, and Reusable (FAIR) [7] but also Emergable (FAIRE) and the goal as a systems-forming factor should be incorporated into the design of the network for epidemics-related projects.

Methods and Algorithms: Our of COVID-19 epidemics [6] is a modification of the classical SIER model. Model diagrams were generated using CellDesigner, a graphical front-end for creating process diagrams of biochemical networks in Systems Biology Markup Language. CellDesigner-generated models were transferred to COPASI (www.copasi.org), which is another Systems Biology Markup Language-compliant program, but with a wider variety of analysis options. The comprehensive dynamic model described in vitro data sets from two independent laboratories. The model and its description is available at FAIRDOMHub [8, 9].

Results: In the presentation, I will analyze the lessons from our COVID-19 modeling, where the model of epidemics [6] was performed along with COVID–19–related clinical data management in the ORCHESTRA EU Horizon 2020 project [https://orchestra-cohort.eu/] focused on:

1. Standardizing and harmonizing COVID-19-related data collection methodologies across participating institutions to ensure consistency and interoperability.
2. Facilitating the sharing of COVID-19 data among researchers and public health authorities across Europe to support collaborative research efforts.
3. Promoting collaboration among researchers from various disciplines to conduct comprehensive studies on COVID-19, including epidemiological modeling, vaccine development, therapeutic interventions, and public health strategies.
4. Providing a platform for rapid data analysis and dissemination of findings to inform decision-making and public health responses to the COVID-19 pandemic [10].

Conclusion: Based on the previous System's Biological experience in COVID-19 modeling and ORCHESTRA clinical data management, I will present the next steps, which could facilitate our preparations for the next potential epidemics.

References

1. Westerhoff H.V., Kolodkin A.N., Conradie R., Wilkinson S.J., Bruggeman F.J., Krab K., Schuppen van J.H., Hardin H., Bakker B.M., Moné M.J., Rybakova K., Eijken M., Leeuwen H.J.P., Snoep J.L. Systems biology towards life in silico: mathematics of the control of living cells. *J Math Biol*. 2009;58(1-2):7-34. doi 10.1007/s00285-008-0160-8
2. Kolodkin A.N., Boogerd F.C., Plant N., Bruggeman F.J., Goncharuk V., Lunshof J., Moreno-Sanchez R., Yilmaz N., Bakker B.M., Snoep J.L., Balling R., Westerhoff H.V. Emergence of the silicon human and network targeting drugs. *Eur J Pharm Sci*. 2012;46(4):190-197. doi 10.1016/j.ejps.2011.06.006
3. Kolodkin A.N., Sahin N., Phillips A., Hood S.R., Bruggeman F.J., Westerhoff H.V., Plant N. Optimization of stress response through the nuclear receptor-mediated cortisol signalling network. *Nat Commun.* 2013;4:1792. doi 10.1038/ncomms2799
4. Kolodkin A.N., Prasad Sharma R., Westerhoff H.V., Colangelo A.M., Ignatenko A., Martorana F., Jennen D., Briede J.J., Brady N., Barberis M., Mondeel T.D.G.A., Papa M., Kumar V., Peters B., Skupin A., Alberghina L., Balling R., Westerhoff H.V. ROS networks: designs, aging, Parkinson's disease and precision therapies. *NPJ Syst Biol Appl*. 2020;6(1):34. doi 10.1038/s41540-020-00150-w
5. Kolodkin A.N., Bruggeman F.J., Plant N., Moné M.J., Bakker B.M., Campbell M.J., van Leeuwen J.P.T.M., Carlberg C., Snoep J.L., Westerhoff H.V. Design principles of nuclear receptor signaling: how complex networking improves signal transduction. *Mol Syst Biol*. 2010;6:446. doi 10.1038/msb.2010.102
6. Westerhoff H.V., Kolodkin A.N. Advice from a systems-biology model of the corona epidemics. *NPJ Syst Biol Appl.* 2020;6(1):18. doi 10.1038/s41540-020-0138-8
7. Santos M.D.V., Anton M., Szomolay B., Kolodkin A. et al. Systems Biology in ELIXIR: modelling in the spotlight. *F1000Res*. 2022;11:ELIXIR-1265. doi 10.12688/f1000research.126734.1
8. Kolodkin A., Westerhoff H.V. Construction of differential equation model to describe COVID-19 epidemics. *FAIRDOMHub.* 2020. https://fairdomhub.org/investigations/372
9. Westerhoff H., Kolodkin A. Comprehensive model for COVID-19 coused by SARS-CoV-2. *FAIRDOMHub*. 2020. doi 10.15490/FAIRDOMHUB.1.MODEL.693.1
10. Tsurkalenko O., Bulaev D., O'Sullivan M.P., Snoeck C., Ghosh S., Kolodkin A. et al. Creation of a pandemic memory by tracing COVID-19 infections and immunity in Luxembourg (CON-VINCE). *BMC Infect Dis*. 2024;24(1):179. doi 10.1186/s12879-024-09055-z

doi 10.18699/bgrs2024-2.1-20

Modeling iron metabolism in patients with post-COVID syndrome

Melchenko N.I.[1]*, Akberdin I.R.[2]

[1] *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

[2] *Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia*

* *n.melchenko@g.nsu.ru*

Key words: mathematical model; iron metabolism; post-COVID syndrome; COVID-19

Motivation and Aim: Most patients who have had SARS-CoV-2 infection suffer from nonspecific symptoms for several months, such as fever, headache, attention disorder, hair loss, etc. Analysis of publications and original experimental data shows that some of these symptoms may be related to impaired iron metabolism, since such patients maintain high concentrations of hepcidin and ferritin for a long time and experience iron deficiency [1, 2]. The construction of a mathematical model describing iron metabolism in patients who have suffered the SARS-CoV-2 infection will help one to identify the key mechanisms and limiting factors underlying long-term disturbances in iron metabolism.
Methods and Algorithms: The model is based on previously published models [3, 4] and constructed using ordinary differential equations (ODEs) with a modular approach. BioUML [5] is used as the software for model construction, simulation and analysis. It enables to develop the model according to state-of-the-art systems biology standards, such as SBGN and SBML [6, 7]. In the reproduced Schirm's model we use the proliferation function and Z-function. The latter is used to describe how erythropoietin effects on amplification and transition times in erythroid cells.
Results: We analyzed a number of existing models describing iron metabolism and transport among different human tissues and organs and selected the most promising ones for our research aim. We have completely reproduced the structure of Mitchell's model [3] by modular approach implemented in BioUML (Fig. 1), as well as all the simulation results obtained in the original study (Fig. 2). Furthermore, we have reproduced the Schirm's model structure [4], describing erythropoiesis and the main components of iron metabolism in details, and have already received simulation dynamics describing Z-functions of the amplifications in compartments: burst forming unit erythroid (BE), colony forming unit erythroid (CE) and the transition times in BE and CE.
Conclusion: As an extension of the model of iron metabolism and transport, which takes into account innate and adaptive immune responses, we incorporate into the modular model of iron metabolism and transport modules of the immune response's impact on hepcidin synthesis, iron storage, labile iron concentration, and intestinal iron absorption. The model of iron metabolism in patients with post-Covid syndrome will allow us to better understand the key pathophysiological mechanisms underlying this condition and post-viral syndromes in general. Development of the model that includes two complex systems such as the immune response and iron metabolism can facilitate the further development of more complex models that can describe the functioning of several human organs and systems.

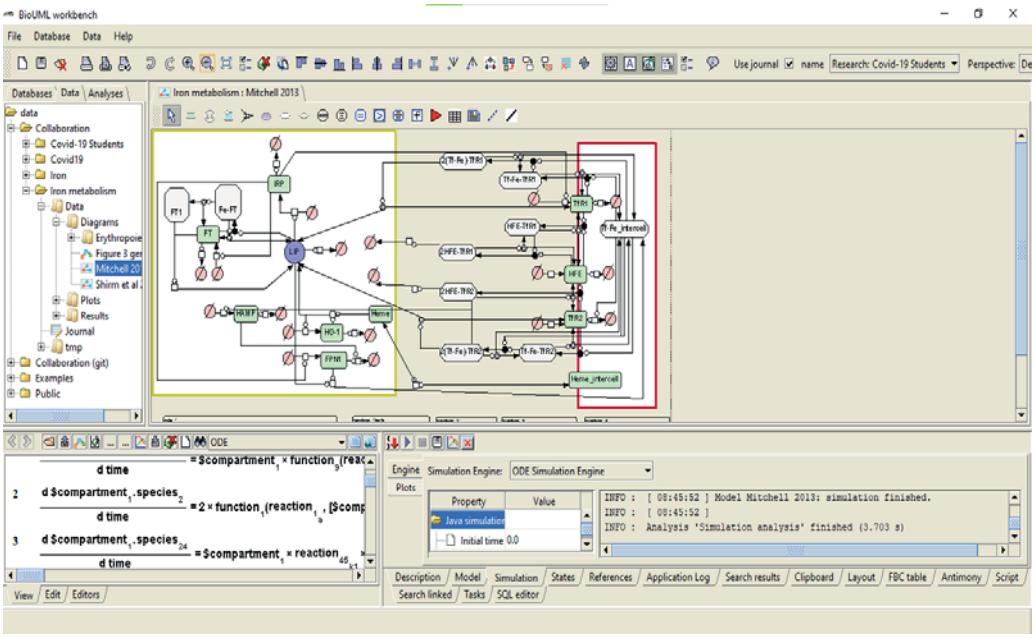


Fig. 1. The modular model of liver iron metabolism reconstructed and analyzed in BioUML from [3]. The compartment with yellow boundary represents the hepatocyte, while the compartment with pink boundary designates plasma. Species overlayed on the compartment boundaries indicate membrane-associated species

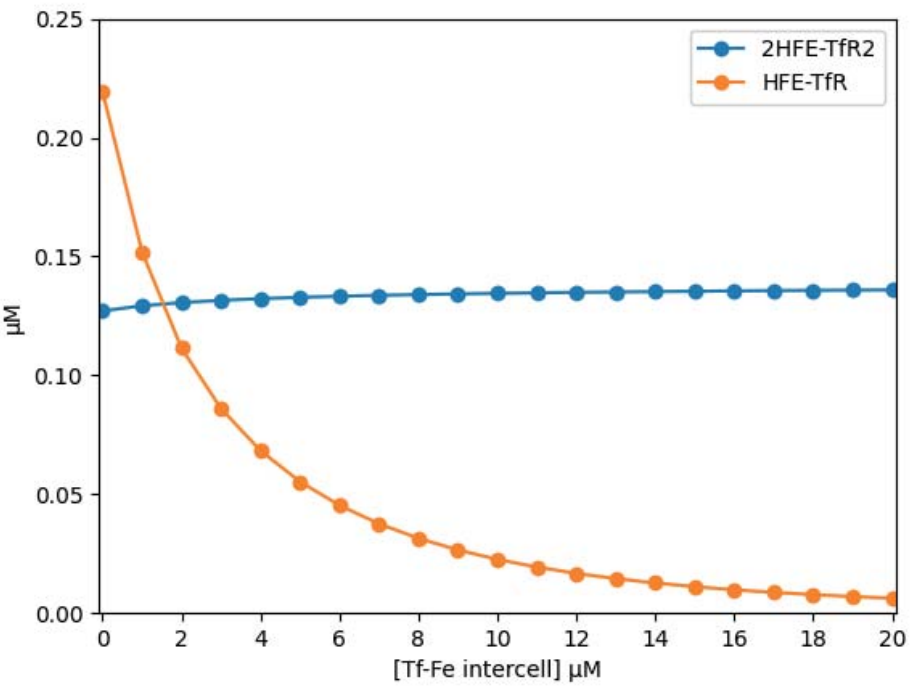


Fig. 2. Reproduced simulation of the steady-state concentrations of metabolites in response to increasing serum Tf-Fe

References

1. Taneri P.E. et al. Anemia and iron metabolism in COVID-19: a systematic review and meta-analysis. *Eur J Epidemiol*. 2020;35(8):763-773. doi 10.1007/s10654-020-00678-5
2. Sonnweber T. et al. The Impact of Iron Dyshomeostasis and Anaemia on Long-Term Pulmonary Recovery and Persisting Symptom Burden after COVID-19: A Prospective Observational Cohort Study. *Metabolites*. 2022;12(6):546. doi 10.3390/metabo12060546
3. Mitchell S., Mendes P.A Computational Model of Liver Iron Metabolism. *PLoS Comput Biol*. 2013;9(11):e1003299. doi 10.1371/journal.pcbi.1003299
4. Schirm S., Scholz M. A biomathematical model of human erythropoiesis and iron metabolism. *Sci Rep*. 2020;10:8602. doi 10.1038/s41598-020-65313-5
5. Kolpakov F. et al. BioUML-towards a universal research platform. *Nucleic Acids Res*. 2022;50(1):124-131. doi 10.1093/nar/gkac286
6. Hucka M. et al. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core Release 2. *J Integr Bioinform*. 2019:16(2):20190021. doi 10.1515/jib-2019-0021
7. Novère N., Hucka M., Mi H. et al. The Systems Biology Graphical Notation. *Nat Biotechnol*. 2009;27:735-741. doi 10.1038/nbt.1558

doi 10.18699/bgrs2024-2.1-21

Modular modeling of human skeletal muscle adaptation to different types of physical exercise considering molecular genetic interactions between metabolism and gene expression regulation

Melikhova E.V.[1]*, Vertyshev A.Yu.[2], Kolpakov F.A.[1], Akberdin I.R.[1]

[1] *Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia*

[2] *JSC "Sites-Tsentr", Moscow, Russia*

* *k_melihovaaa@mail.ru*

Key words: mathematical modeling; modular approach; skeletal muscle; physical exercise; BioUML

Motivation and Aim: Nowadays there is a growing society's interest in research studies and the development of technologies related to health preservation. It is known that physical activity plays an important role in the development and maintenance of a functionally healthy state not only of individual organs, but also of the entire human body. Regular physical exercise affects almost all cell types, tissues and organs. The cumulative effect of each exercise bout improves the health and performance depending on the tissue [1]. However, achievement of the best outcomes in the training process is feasible with a deep insight of not only physiological, but also molecular-genetic mechanisms underlying the body's response to physical exertion.

Muscle tissue is one of the four main tissues present in the human body. Skeletal muscles constitute approximately 40 % of the total body mass in a healthy adult and serve a multitude of functions, from movement to maintaining metabolic balance [2, 3]. The functions of movement and maintaining body posture are provided by the unique ability of muscles to contract.

During physical exertion, a series of changes occur in skeletal muscle cells at the transcriptomic, proteomic and metabolomic levels [4]. Furthermore, muscles undergo structural and functional changes with regular exercise that enables them to work more efficiently. Enzyme synthesis plays a key role in the adaptation of muscles to physical activity. Various methods can be employed to assess the regulation of enzyme synthesis during adaptation to physical exercise, such as measuring enzyme activity in muscle tissue, measuring the expression levels of genes encoding these enzymes, or studying changes in metabolite levels in muscle tissue. However, *in vivo* studies have several limitations, including the absence of implementation at the level of individual cells, technical measurement complexities, labor intensiveness, and costliness.

A reconstruction of the mathematical model describing skeletal muscle adaptation to various types of the physical activity avoids these constraints, which makes it possible to analyze the interaction between molecular components in a systematic way. The existing model of skeletal muscle adaptation, developed by Akberdin et al. [5], has a hierarchical structure and includes physiological and cellular levels. At the first level the interaction of muscle fibers with the circulatory blood system is described, while the second one comprises the processes occurring in metabolism, signaling pathways and gene expression. The model links metabolism, Ca^{2+} and AMPK-dependent signal trans-

duction pathways and regulation of gene expression in human skeletal muscles. However, the modular model uses the phenomenological linear function that affects the rate of enzymatic reactions during the exercise and does not consider concentration dynamics of enzymes catalyzing reactions of the central metabolism in fiber cells assuming that their concentrations are constant during the exercise mode.

An analysis of publications and omics data [6, 7] allowed us to determine what changes are observed in the expression of genes encoding the enzyme of central metabolism and the synthesis of these proteins. Using the GTRD database [8], transcription factors associated with genes encoding central metabolic enzymes have been identified. The data provides the qualitative and quantitative basis to consider such molecular processes as transcription and translation of enzymes in the muscle cell taking into account expression and action of transcription factors orchestrating their mRNA synthesis.

Thus, the aim of this study is an extension of the modular model structure to adequately describe changes in enzyme concentrations during an exercise and recovery.

Methods and Algorithms: The modeling was performed using the tool sets of the BioUML software, which is developed for mathematical modeling of life systems and analysis of various types of biological data [9]. The platform supports main standards for visualization and simulation of biological systems, a number of numerical methods, and integration with Jupyter Notebook to ensure reproducibility of research results.

Visual modeling implemented in the platform provides the representation of the studied systems as graphical diagrams, which greatly facilitates the understanding of complex biological processes. Each component of the diagram corresponds to a specific object (variable, metabolite, reaction, etc.). Based on the defined properties of the components and their interconnections, BioUML automatically generates Java code used for the model compilation and simulation.

The modular approach in BioUML implies the representation of the system under study as a set of interconnected subsystems. Each module is a mathematical model that can be considered and simulated independently. The combination and integration of modules constitute a complex, comprehensive model of the entire system. The modularity principle enables the development and adaptation of the model to changing needs without the requirement of a complete rebuild of the underlying architecture. Additionally, this approach allows researchers to select, use, and study only those modules that are relevant to specific tasks.

Results: As a result of the study, a modular model describing the adaptation of skeletal muscle to different types of physical exercise was expanded by considering processes both regulating expression of genes encoding enzymes and protein synthesis. The extension of the transcription factor module, along with the addition of the transcription and translation modules for central metabolism's enzymes facilitates more precise and detailed description of the molecular mechanisms of human skeletal muscle adaptation to various types of physical stress. This opens up the opportunity for a design of optimal training strategies to achieve specific goals for individuals.

Conclusion: The modular model describing the human skeletal muscle adaptation to different modes of the physical exercise has been extended by additional modules for transcription and translation processes required for adequate simulation of changes in enzyme concentration. The next stage of the study is to verify the quality of the model by comparing simulation results to accumulated experimental data for different types of physical exertion.

Funding: The study was financially supported by the Russian Science Foundation (project No. 24-14-20031, https://www.rscf.ru/en/project/24-14-20031/).

References

1. Ashcroft S.P., Stocks B., Egan B., Zierath J.R. Exercise induces tissue-specific adaptations to enhance cardiometabolic health. *Cell Metab.* 2024;36(2):278-300. doi 10.1016/j.cmet.2023.12.008
2. Frontera W.R., Ochala J. Skeletal muscle: a brief review of structure and function. *Calcif Tissue Int.* 2015;96(3):183-195. doi 10.1007/s00223-014-9915-y
3. Weskamp K., Olwin B.B., Parker R. Post-Transcriptional Regulation in Skeletal Muscle Development, Repair, and Disease. *Trends Mol Med.* 2021;27(5):469-481. doi 10.1016/j.molmed.2020.12.002
4. Egan B., Zierath J.R. Exercise metabolism and the molecular regulation of skeletal muscle adaptation. *Cell Metab.* 2013;17(2):162-184. doi 10.1016/j.cmet.2012.12.012
5. Akberdin I.R., Kiselev I.N., Pintus S.S., Sharipov R.N., Vertyshev A.Y., Vinogradova O.L., Popov D.V., Kolpakov F.A. A Modular Mathematical Model of Exercise-Induced Changes in Metabolism, Signaling, and Gene Expression in Human Skeletal Muscle. *Int J Mol Sci.* 2021;22(19):10353. doi 10.3390/ijms221910353
6. Popov D.V., Makhnovskii P.A., Shagimardanova E.I., Gazizova G.R., Lysenko E.A., Gusev O.A., Vinogradova O.L. Contractile activity-specific transcriptome response to acute endurance exercise and training in human skeletal muscle. *Am J Physiol Endocrinol Metab.* 2019;316(4):E605-E614. doi 10.1152/ajpendo.00449.2018
7. Makhnovskii P.A., Zgoda V.G., Bokov R.O. et al. Regulation of Proteins in Human Skeletal Muscle: The Role of Transcription. *Sci Rep.* 2020;10:3514. doi 10.1038/s41598-020-60578-2
8. Kolmykov S., Yevshin I., Kulyashov M., Sharipov R., Kondrakhin Y., Makeev V.J., Kulakovskiy I.V., Kel A., Kolpakov F. GTRD: an integrated view of transcription regulation. *Nucleic Acids Res.* 2021;49(D1):D104-D111. doi 10.1093/nar/gkaa1057
9. Kolpakov F., Akberdin I., Kashapov T., Kiselev L., Kolmykov S., Kondrakhin Y., Kutumova E., Mandrik N., Pintus S., Ryabova A., Sharipov R., Yevshin I., Kel A. BioUML: an integrated environment for systems biology and collaborative analysis of biomedical data. *Nucleic Acids Res.* 2019;47(W1):W225-W233. doi 10.1093/nar/gkz440

doi 10.18699/bgrs2024-2.1-22

AMPA receptors endocytosis pathways involve into complexity regulatory network of synaptic plasticity

Proskura A.L., Ratushnyak A.S.*
Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia
** ratushniak.alex@gmail.com*

Key words: endocytosis; AMPA receptors; dendritic spine; hippocampus

Motivation and Aim: The dendritic microprotrusions – dendritic spines – are the postsynaptic part of exciting synapses. Ionotropic glutamate receptors (NMDAR (The N-methyl-D-aspartate receptor), AMPAR (The α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor) are key components of exciting synapses into a brain. Through structural proteins of postsynaptic density, they form various protein macrocomplexes [1]. It mediated of intermolecular rearrangements and form the functional regulatory networks. In addition, the receptors of various hormones, in particular insulin, growth and neurotrophic factors, as well as the receptors of other mediator systems are present in the synaptic contact zone. However, the molecular mechanisms of links synaptic signalling pathways to the AMPARs endocytosis machinery remains elusive.
Methods and Algorithms: The interactome of hippocampal dendritic spines was reconstructed (technology GeneNet was used (РОСПАТЕНТ № 990006 от 15/02/1999)). Synaptic processes of the adult hippocampus have been reconstructed on the basis of GeneNet information (http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/viewer) [2].
Results: The AMPAR is subject to functionally distinct constitutive and regulated clathrin-dependent endocytosis, contributing to various forms of synaptic plasticity. On induction of synaptic plasticity, internalized AMPA receptors undergo endosomal sorting and cycle through early endosomes and recycling endosomes back to the plasma membrane (model for long-term potentiation) or target for degradation to the lysosomes (model for long-term depression) [3]. The reception of the basic mediator signal – glutamat – is carried out by NMDAR macrocomplexes. NMDAR activation leads to clathrin-dependent endocytosis of postsynaptic AMPA receptors. A key event is the groundwork of second messengers pools. P4 → PIP2 (PtdIns(4,5)P2) catalyzated Phosphoinositide 5-kinase (PI5K), PIP2 → PIP3 (PtdIns(3,4,5)P3) mediated Phosphoinositide 3-kinase (PI3K). Ca^{2+} influx through the NMDA receptor activates calcineurin and protein phosphatase 1 to dephosphorylate and activate phosphatidylinositol 4-phosphate 5-kinase (PIP5K), the major phosphatidylinositol 4,5-bisphosphate ($PI(4,5)P_2$)-producing enzyme in the brain [4]. PtdIns(4,5)P2 is a substrate for hydrolysis by phospholipase C (PLC). The products of the PLC catalyzation of PIP2 are inositol 1,4,5-trisphosphate (InsP3; IP3) and diacylglycerol (DAG), both of which function as second messengers. In this cascade, DAG remains on the cell membrane and activates the signal cascade by activating protein kinase C (PKC). GluR2-AMPAR, which provide the basic neurotransmission are move from the synaptic contacts zone after phosphorylation by PKC on S880 of GluR2 subunit cytoplasmic domain [5]. Metabotropic glutamate receptors I class, mGluR1 and mGluR5, involve in regulate of synaptic plasticity processes thru activate PLC [6].

PIP5K activation alternative pathway in the dendritic spine interactome during NMDAR-mediated synaptic plasticity was our reconstructed: ARF6 → PI(4)P5K alpha → PI(4,5)P2 (PIP2). PI(4)P5Kalpha is a downstream effector of ARF6 and when ARF6 is activated, it triggers recruitment of a diverse but interactive set of signaling molecules into sites of active cytoskeletal and membrane rearrangement, involved in on receptor-mediated endocytosis plasma, a membrane recycling pathway [7–9]. ARF6 activators, such as EFA6A, BRAG, are highly concentrated in the postsynaptic density fraction and formed a protein complex with NMDAR [10, 11]. BRAG2 regulates Arf6-dependent endocytosis of AMPARs through the direct interaction decrease in their synaptic expression in hippocampal neurons that kay during the hippocampal long-term depression [12]. Insulin is the most abundant peptidergic hormone secreted by the pancreatic islets of Langerhans and plays an important role in organic metabolism. The insulin receptor signaling in the brain have been various functions for normal neurophysiology, and a dysregulation of insulin secretion or insulin receptor signaling has been reported in serious mental illnesses [13]. Insulin has also been shown to regulate the endocytosis of synaptic AMPA receptors [14]. Insulin effects mediated of PIP3 accumulation, that provides the prolonged of PKC activity and that contribute to GluR2-AMPAR endocytosis. It is important to emphasize that for PIP3 accumulation it is necessary to turn off the main antagonist of PI3K – PTEN (phosphatase and tensin homolog deleted on chromosome 10). The reactive oxygen species (ROS) turn off PTEN what needed to realize the effect of insulin [15].

Conclusion: AMPAR endocytosis are key event during synaptic plasticity regulation. There are different and/or alternative pathways for triggering endocytosis of synaptic receptors, that is under the control of a complexity regulatory network in the dendritic spine interactome.

Funding: The research was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation for Federal Research Center for Information and Computational Technologies.

References

1. Proskura A.L., Ratushnyak A.S., Vechkapova S.O., Zapara T.A. Synapse as a multi-component and multi-level information system. *Stud Comput Intell*. 2018;736:186-192
2. Proskura A.L., Malakhin I.A., Turnaev I.I., Suslov V.V., Zapara T.A., Ratushnyak A.S. Intermolecular interactions in the functional systems of neuron. *Vavilov Zh Genet Selekts*. 2013;17:620-628
3. van der Sluijs P., Hoogenraad C.C. New insights in endosomal dynamics and AMPA receptor trafficking. *Semin Cell Dev Biol*. 2011;22(5):499-505. doi 10.1016/j.semcdb.2011.06.008
4. Unoki T., Matsuda S., Kakegawa W., Van N.T., Kohda K., Suzuki A., Funakoshi Y., Hasegawa H., Yuzaki M., Kanaho Y. NMDA receptor-mediated PIP5K activation to produce $PI(4,5)P_2$ is essential for AMPA receptor endocytosis during LTD. *Neuron*. 2012;73(1):135-148. doi 10.1016/j.neuron.2011.09.034
5. Seidenman K.J., Steinberg J.P., Huganir R., Malinow R. Glutamate receptor subunit 2 Serine 880 phosphorylation modulates synaptic transmission and mediates plasticity in CA1 pyramidal cells. *J Neurosci*. 2003;23(27):9220-9228. doi 10.1523/JNEUROSCI.23-27-09220.2003
6. Neyman S., Manahan-Vaughan D. Metabotropic glutamate receptor 1 (mGluR1) and 5 (mGluR5) regulate late phases of LTP and LTD in the hippocampal CA1 region in vitro. *Eur J Neurosci*. 2008;27(6):1345-1352. doi 10.1111/j.1460-9568.2008.06109.x
7. Honda A., Nogami M., Yokozeki T., Yamazaki M., Nakamura H., Watanabe H., Kawamoto K., Nakayama K., Morris A.J., Frohman M.A., Kanaho Y. Phosphatidylinositol 4-phosphate 5-kinase alpha is a downstream effector of the small G protein ARF6 in membrane ruffle formation. *Cell*. 1999;99(5):521-532. doi 10.1016/s0092-8674(00)81540-8
8. D'Souza-Schorey C., Li G., Colombo M.I., Stahl P.D. A regulatory role for ARF6 in receptor-mediated endocytosis. *Science*. 1995;267(5201):1175-1178. doi 10.1126/science.7855600

9. Radhakrishna H., Donaldson J.G. ADP-ribosylation factor 6 regulates a novel plasma membrane recycling pathway. *J Cell Biol*. 1997;139(1):49-61. doi 10.1083/jcb.139.1.49
10. Sakagami H., Honma T., Sukegawa J., Owada Y., Yanagisawa T., Kondo H. Somatodendritic localization of EFA6A, a guanine nucleotide exchange factor for ADP-ribosylation factor 6, and its possible interaction with alpha-actinin in dendritic spines. *Eur J Neurosci*. 2007;25(3):618-628. doi 10.1111/j.1460-9568.2007.05345.x
11. Scholz R., Berberich S., Rathgeber L., Kolleker A., Köhr G., Kornau H-C. AMPA receptor signaling through BRAG2 and Arf6 critical for long-term synaptic depression. *Neuron*. 2010;66(5):768-780. doi 10.1016/j.neuron.2010.05.003
12. Fukaya M., Sugawara T., Hara Y., Itakura M., Watanabe M., Sakagami H. BRAG2a Mediates mGluR-Dependent AMPA Receptor Internalization at Excitatory Postsynapses through the Interaction with PSD-95 and Endophilin 3. *J Neurosci*. 2020;40(22):4277-4296. doi 10.1523/JNEUROSCI.1645-19.2020
13. Grillo C.A. et al. Hippocampal insulin resistance impairs spatial learning and synaptic plasticity. *Diabetes*. 2015;64(11):3927-3936. doi 10.2337/db15-0596
14. Huang C.C. et al. The role of insulin receptor signaling in synaptic plasticity and cognitive function. *Chang Gung Med J*. 2010;33(2):115-125
15. Seo J.H., Ahn Y., Lee S-R., Yeo C.Y., Hur K.C. The major target of the endogenously generated reactive oxygen species in response to insulin stimulation is phosphatase and tensin homolog and not phosphoinositide-3 kinase (PI-3 kinase) in the PI-3 kinase/Akt pathway. *Mol Biol Cell*. 2005;16(1):348-357. doi 10.1091/mbc.e04-05-0369

doi 10.18699/bgrs2024-2.1-23

Identification of the molecular mechanisms underlying the negative effects of nitrogen fertilizers on resistance to sheath blight in rice

Volyanskaya A.R.[1], Antropova E.A.[1], Demenkov P.S.[1, 2], Ivanisenko V.A.[1, 2]*

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *salix@bionet.nsc.ru*

Key words: *Oryza sativa*; RNA-seq; differentially expressed genes; nitrogen fertilization

Motivation and Aim: Rice (*Oryza sativa* L.) is considered the most valuable crop in the world, as half of the world's humanity uses it for food every day. Sheath blight disease, caused by fungus *Rhizoctonia solani* Kühn causes serious damage to rice yields, leading to losses of up to 50 % of the yield [2]. It is well known that the use of fertilizers increases the productivity of agricultural plants. However, in addition to the positive effect, nitrogen fertilization has been shown to worsen the resistance of rice to sheath blight [3]. The molecular mechanisms of this effect are still not understood and require more thorough research.

Methods and Algorithms: Publicly available transcriptomic data (RNA-seq) on the response to nitrogen fertilizers and sheath blight infection were analyzed to identify differentially expressed genes. To reconstruct and analyze gene networks, the SmartSrop cognitive platform, developed by us, was used, which was created to study the genotype-phenotype-environment relationships of agriculturally valuable crops rice and wheat using artificial intelligence methods and text analysis.

Results: We analyzed transcriptomic data on the response of rice to *Rhizoctonia solani* infection, identified a group of differentially expressed genes, and reconstructed the corresponding gene network. Similar work was carried out on publicly available data on the effects of nitrogen fertilizers on the rice transcriptome. Next, the resulting gene networks were analyzed in the SmartSrop cognitive platform to search for mechanisms of the negative effect of nitrogen fertilizers on resistance to sheath blight.

Conclusion: The data obtained allow us to make assumptions about the mechanisms underlying the negative effect of fertilizers on resistance to sheath blight.

Funding: The study is supported by the Russian-Chinese grant from the Russian Science Foundation No. 23-44-00030.

References

1. Bin Rahman A.N.M.R., Zhang J. Trends in rice research: 2030 and beyond. *Food Energy Secur*. 202;12(2):e390. doi 10.1002/fes3.390
2. Senapati M., Tiwari A., Sharma N., Chandra P., Bashyal B.M., Ellur R.K., Bhowmick P.K., Bollinedi H., Vinod K.K., Singh A.K., Krishnan S.G. Rhizoctonia solani Kühn Pathophysiology: Status and Prospects of Sheath Blight Disease Management in Rice. *Front Plant Sci*. 2022;13:881116. doi 10.3389/fpls.2022.881116
3. Tang Q., Peng S., Buresh R., Zou Y., Castilla N.P., Mew T.W., Zhong X. Rice varietal difference in sheath blight development and its association with yield loss at different levels of N fertilization. *Field Crops Res*. 2007;102:219-227. doi 10.1016/j.fcr.2007.04.005

doi 10.18699/bgrs2024-2.1-24

Mathematical modeling of AMPK/mTOR signaling system uncovers the causes in differences of hypothalamic neuropeptide expression regulating feeding behavior in slow- and fast-growing chickens

Volyanskaya A.R.[1]*, Osik N.A.[2], Yanshole L.V.[2], Tsentalovich Yu.P.[2], Gusev O.A.[3, 4, 5], Kolpakov F.A.[6], Akberdin I.R.[6]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *International Tomography Center, SB RAS, Novosibirsk, Russia*

[3] *Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[4] *Life Improvement by Future Technologies (LIFT) Center, Moscow, Russia*

[5] *Intractable Disease Research Center, Juntendo University, Tokyo, Japan*

[6] *Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia*

* *a.volianskaia@alumni.nsu.ru*

Key words: mathematical modeling; poultry; feed intake regulation; AMPK/mTOR signaling pathway

Motivation and Aim: Breeding of chickens for high growth rates (HGR) and high body weight has led to changes in the feeding behavior (FB) of meat chicken lines that are unable to adequately regulate feed intake commensurate with their energy needs. This is accompanied by the accumulation of adipose tissue, undesirable body composition, as well as a number of health problems [1]. Thus, for the commercial development of chicken meat lines, a detailed study of the mechanisms underlying the regulation of FB and energy balance of poultry is necessary. The regulation of FB in chickens, like mammals, is mainly carried out by neural networks in the hypothalamus. An important role in the regulation of FB is assigned to specific neuropeptides expressed in the nuclei of the hypothalamus. Depending on their effect on appetite, hypothalamic neuropeptides (HN) have been divided into two groups: orexigenic (OHN) and anorexigenic (AHN). The group of OHNs that increase appetite and create a state of hunger includes neuropeptide Y (NPY) and agouti-related protein (AgRP). AHNs, which include alpha-melanocyte-stimulating hormone (α-MSH), on the contrary, suppress appetite and create a state of satiety. A key role in the regulation of FB is played by the signaling of the hormones insulin and leptin through their influence on the expression of OHNs and AHNs [2]. Hormone signaling through hypothalamic neural networks is closely linked to the adenosine monophosphate-activated protein kinase/mammalian target of rapamycin (AMPK/mTOR) signaling pathway [3]. It has been shown that the expression levels of HNs that mediate the regulation of FB differ greatly between chickens with HGR and slow growth rates (SGR). In meat breeds of chicken, in contrast to SGR chickens, a decrease in the expression level of the proopiomelanocortin *(POMC)* gene, encoding the AHN α-MSH, is observed with age, which may be one of the reasons for excessive feed consumption in broilers [4, 5]. This observation may be based on a difference in signaling of the transcription factor (TF) FOXO1, which is a transcriptional activator of the *POMC* gene. The multiple interactions between AMPK/mTOR and hormone-mediated

signaling pathways, with subsequent regulation of HNs expression, result in complex dynamic behavior of this signaling-metabolic system that is not intuitively predictable. Thus, the application of a mathematical modeling approach, including the development of detailed mechanistic and modular models, is critical to further study the molecular mechanisms and their influence on FB and energy balance in chickens.

Methods and Algorithms: Quantitative metabolomic analysis of chicken tissues was carried out using the NMR method. Metabolomic extracts of chicken heart, liver, kidney, breast and leg muscle were prepared as described in [6, 7], and then NMR spectra of the extracts were obtained using an NMR spectrometer AVANCE III HD 700 MHz (Bruker BioSpin, Germany). For each tissue type, the metabolomic analysis was performed for 12 samples taken from SGR chickens and 12 samples from FGR. Quantification of NMR data was performed by integrating the NMR signals relative to an internal standard. Mathematical models describing the mechanisms of FB regulation in chicken hypothalamic cells with SGR and HGR were reconstructed in the open access BioUML software package [8]. The structure of the developed model is based on the Sonntag model [9], which considers insulin and amino acid signaling as regulators of the AMPK/mTOR signaling system. The signaling pathway of leptin is implemented based on literature data [10] and the mechanisms described in the Yamada model [11]. The structure of the cytosolic and mitochondrial modules, as well as the transport modules, which describe metabolic processes including glycolysis, the tricarboxylic acid cycle, oxidation and biosynthesis of fatty acids, was taken from the model [12], which considers metabolic processes in human skeletal muscles. To expand mathematical models, a cell nucleus module was created, which describes the processes of genetic regulation of the expression of *NPY*, *AgRP* and *POMC* genes encoding HNs.

Results: We determined the concentrations (in units of nmol per gram of a tissue) of 52 metabolites in the chicken breast muscle, 57 metabolites in the leg muscle, 66 metabolites in the heart, 75 metabolites in the kidney, and 74 metabolites in the liver. However, chemometric analysis did not reveal metabolites with statistically significant differences between SGR and FGR chickens. It is possible that the metabolomic differences are too small compared to the variation within each group, and a significantly larger sample size must be used to detect these differences. A composite mathematical model was built, consisting of 6 modules: a cytosol module, which describes the AMPK/mTOR signaling pathway; a nuclear module containing regulation of the expression of HNs; a cytosol and mitochondria modules describing central metabolic processes considering changes in the concentration of metabolites measured by the NMR method, and three transport modules. As part of the modeling of free-feeding conditions for chickens, based on experimental data, approximating functions were proposed that describe changes in leptin concentration in SGR and HGR chickens depending on the age [13]. Mathematical models simulating free-feeding conditions have been fitted to hypothalamic NPY and POMC mRNA levels in SGR and HGR chickens [14, 15]. The structural differences in the regulation of the expression of these genes were based on the experimentally identified difference between two breeds of chickens: in FGR chickens, the TF FOXO1 was not phosphorylated in the insulin signaling pathway [4, 5]. This mechanism of transcriptional regulation made it possible to obtain differences in the dynamics of expression of the *POMC* and *NPY* genes in accordance with experimental data.

Conclusion: Mathematical models were constructed to describe the regulation of feeding behavior in the cells of the hypothalamus in chickens with a SGR and HGR. The basis

for the structural difference between the models was the observed difference in the signaling of the TF FOXO1, which regulates the expression of HNs. Modeling of the dynamics of expression of the concentrations of the HNs NPY and POMC was carried out under conditions of free feeding of chickens, proposed on the basis of the functions of changes in leptin concentrations in chickens with a SGR and HGR. Structural differences in the FOXO1 signaling pathway mediated differences in the dynamics of POMC and NPY concentrations. Broilers showed an increase in POMC concentrations with age, while chickens with a SGR showed a decrease. This mechanism may explain the difference in FB of meat chicken breeds.

Funding: This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2021-1344).

References

1. Richards M.P., Proszkowiec-Weglarz M. Mechanisms regulating feed intake, energy expenditure, and body weight in poultry. *J Poult Sci*. 2007;86(7):1478-1490
2. Boswell T. Regulation of energy balance in birds by the neuroendocrine hypothalamus. *J Poult Sci*. 2005;42(3):161-181
3. Proszkowiec-Weglarz M. et al. Characterization of the AMP-activated protein kinase pathway in chickens. *Comp Biochem Physiol B: Biochem Mol Biol*. 2006;143(1):92-106
4. Saneyasu T. et al. Age-dependent changes in the mRNA levels of neuropeptide Y, proopiomelanocortin, and corticotropin-releasing factor in the hypothalamus in growing broiler chicks. *J Poult Sci*. 2013;50(4):364-369
5. Honda K. et al. Correlation analysis of hypothalamic mRNA levels of appetite regulatory neuropeptides and several metabolic parameters in 28-day-old layer chickens. *Anim Sci J*. 2015;86(5):517-522
6. Tsentalovich Y.P. et al. Seasonal variations and interspecific differences in metabolomes of freshwater fish tissues: Quantitative metabolomic profiles of lenses and gills. *Metabolites*. 2019;9(11):264
7. Tsentalovich Y.P. et al. Most abundant metabolites in tissues of freshwater fish pike-perch (Sander lucioperca). *Sci Rep*. 2020;10(1):17128
8. Kolpakov F. et al. BioUML – Towards a universal research platform. *Nucleic Acids Res*. 2022;50(W1):W124-W131
9. Sonntag A.G. et al. A modelling–experimental approach reveals insulin receptor substrate (IRS)-dependent regulation of adenosine monosphosphate-dependent kinase (AMPK) by insulin. *FEBS J*. 2012;279(18):3314-3328
10. Barrios-Correa A.A., Estrada J.A., Contreras I. Leptin signaling in the control of metabolism and appetite: lessons from animal models. *J Mol Neurosci*. 2018;66(3):390-402
11. Yamada S. et al. Control mechanism of JAK/STAT signal transduction pathway. *FEBS Lett*. 2003;534(1-3):190-196
12. Akberdin I.R. et al. A modular mathematical model of exercise-induced changes in metabolism, signaling, and gene expression in human skeletal muscle. *Int J Mol Sci*. 2021;22(19):10353
13. Cassy S. et al. Peripheral leptin effect on food intake in young chickens is influenced by age and strain. *Domest Anim Endocrinol*. 2004;27(1):51-61
14. Saneyasu T. et al. Age-dependent changes in the mRNA levels of neuropeptide Y, proopiomelanocortin, and corticotropin-releasing factor in the hypothalamus in growing broiler chicks. *J Poult Sci*. 2013;50(4):364-369
15. Honda K. et al. Correlation analysis of hypothalamic mRNA levels of appetite regulatory neuropeptides and several metabolic parameters in 28-day-old layer chickens. *Anim Sci J*. 2015;86(5):517-522

doi 10.18699/bgrs2024-2.1-25

Retrotransposons expression controls: hubs and networks

Yuditskiy K.I., Samsonova A.A., Kanapin A.A.*

Institute of Translational Biomedicine, Saint Petersburg State University, St. Petersburg, Russia
** akanapin@spbu.ru*

Key words: retrotransposons; transcriptional regulation; gene networks; epigenetics

Motivation and Aim: Over half of the human genome consists of interspersed repeats that result from copy-and-paste events carried out by retrotransposons [1]. These elements are typically kept in check through epigenetic silencing to prevent genome instability. In cancer, changes in the epigenetic landscape can lead to abnormal expression of these genetic elements, which can act as internal mutagens and contribute to clonal evolution and cancer development [2]. Our specific goal is to uncover new regulatory mechanisms, networks, and potentially genomic elements that control the transcriptional activity of retroelements in both cancer and normal cells. By conducting a thorough analysis of high-throughput transcriptional datasets and developing new computational frameworks, we hope to identify potential regulatory mechanisms that control retrotransposon expression in both normal and cancer cells, bridging the gap between seemingly unrelated observations of epigenetic target inhibition and retrotransposon expression.
Methods and Algorithms: The experiments relevant to this study were chosen from a literature search of published RNASeq experiments involving HDAC inhibitors. The publicly available transcriptomic data and corresponding metadata were downloaded from the SRA (Short Read Archive) repository and the Entrez database, respectively. A complete list of data accessions is available in the project database: https://transrt.compbio.ru/db/. The SalmonTE package [3] was used to assess the expression levels of both genes and transposons, which included a total of 688 transposons across 14 different types. The normalized gene and transposon expression data were then used to predict gene regulatory networks using the WGCNA package [4]. Consensus networks were constructed to identify control modules that remained consistent across experiments, and adjacency matrices were created for each experiment using the WGCNA::adjacency method. The obtained minimum and median adjacency matrices were used to calculate the Topological Overlap Matrix (TOM). Subsequently, the dendrogram of co-expression modules was obtained using the TOMs, and the modules were extracted by applying the dynamicTreeCut::cutreeDynamic algorithm to the dendrogram. Additionally, the ARACHNe-AP package [5] was used to identify transcription factors that control retrotransposon expression.
Results: We have successfully constructed distinct gene networks for various experiments and have identified co-expression modules that align with each other. Consensus modules, which remain consistent across cell types, have been determined for two retrotransposon families: LINE1 and HERV. The modules are mainly composed of long non-coding RNAs, making it difficult to perform standard gene set enrichment analyses and functional annotation.
Using ARACHNe-AP algorithms, we conducted a systematic search for transcription factors that control transposon activity. This allowed us to identify specific transcription

factor-transposon pairs and generate lists of factors that exclusively regulate certain classes of transposon elements, as well as general regulators. Based on the results from WGCNA, which revealed two distinct regulatory modules for LINE and ERV transposons, we carried out an enrichment analysis of Reactome and KEGG signaling pathways associated with the transcription factors that exclusively regulate these transposon classes. Notably, we observed a significant difference in the enriched transcription factors for LINE and HERV transposons.

Conclusions: Our findings offer fresh insights into the intricate regulatory networks governing the expression of retrotransposons in both normal and cancer cells. It was observed that transposons from at least two classes, namely LINE and HERV are subject to regulation through distinct mechanisms. Long non-coding RNAs were found to play a significant role in the regulation of both types of transposons. Furthermore, the distinct sets of transcriptional regulators controlling the expression of different mobile elements suggest differing mechanisms governing retrotransposon functions in normal and pathological cells.

Funding: The study is supported by Russian Science Foundation (Grant No. 23-14-00134).

References

1. de Koning A.P., Gu W., Casto T.A., Batze M A., Polloc D.D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 2011;7(12):e1002384. doi 10.1371/journal.pgen.1002384
2. Roulois D., Loo Yau H., Singhania R., Wang Y., Danesh A., Shen S.Y., Han H., Liang G., Jones P.A., Pugh T.J., O'Brien C., De Carvalho D.D. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell.* 2015;162(5):961-973. doi 10.1016/j.cell.2015.07.056
3. Jeong H.H., Yalamanchili H.K., Guo C., Shulman J.M., Liu Z. An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. *Pac Symp Biocomput.* 2018;23:168-179
4. Langfelder P., Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. doi 10.1186/1471-2105-9-559
5. Lachmann A., Giorg, F.M., Lopez G., Califano A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics*. 2016;32(14):2233-2235. doi 10.1093/bioinformatics/btw216

2 Симпозиум «Системная компьютерная биология»
Symposium “Systems computational biology”

doi 10.18699/bgrs2024-2.2-01

Роль фотосинтетических процессов в жизнедеятельности фитопланктона. Математическая модель

Абакумов А., Пак С.*

Институт автоматики и процессов управления ДВО РАН, Владивосток, Россия
** packsa@iacp.dvo.ru*

Ключевые слова: математическая модель; дифференциальные уравнения; фитопланктон; фотосинтез; хлорофилл; фотосинтетическая активная радиация; температура; задача оптимального управления

Мотивация и цель: Фитопланктон является ключевым звеном глобального углеродного цикла. На первый план ставится изучение механизмов образования первичной продукции. Существуют способы расчета первичной продукции от массы фитопланктона с применением ассимиляционного числа [1, 2]. Однако такая схема представляется упрощенной, поскольку не учитывает переменчивость количества фотосинтезирующих веществ в клетках фитопланктона, основным среди которых является хлорофилл. Доля хлорофилла по массе в фитопланктоне меняется в зависимости от состояния клеток и условий обитания. Массовую долю хлорофилла в фитопланктоне принято называть хлорофилльной квотой [3]. От ее величины во многом зависят фотосинтетические характеристики планктонного сообщества. Все биохимические процессы, влияющие на динамику биомассы, осуществляются при участии энергоемких веществ, макроэргов, которые присутствуют в живых клетках преимущественно в форме аденозинтрифосфата (АТФ). Их концентрация на сегодняшний день может измеряться непосредственно. Современные методы исследований позволяют судить об интенсивности продукционных процессов на основании подобных измерений. Предыдущий этап исследований заключался в разработке математических моделей, в которых продукция фитопланктона рассчитывается с учетом изменчивости хлорофилльной квоты, а также концентрации энергоемких веществ. Исследование свойств моделей направлено на оптимизацию методов решения с целью расширить область применения модели к реальным биологическим системам и упрощение процесса сбора данных.
Методы и алгоритмы: Разработка моделей основана на использовании модели Друпа, которая, в отличие от общепринятой кинетики Моно, является более гибкой и способна описать продолжение роста фитопланктона в течение некоторого времени после того, как минеральное питание перестает поступать в систему. Такая же форма зависимости была заимствована для описания роста биомассы на темновой стадии фотосинтеза, когда иссякает освещение как необходимый ресурс. Исследование свойств моделей включает проверку устойчивости равновесных решений по Ляпунову, а также построение фазовых портретов в пространствах ведущих параметров модельных систем. Исследование поведения модели на аттракторе дает более полное представление о ее свойствах. Решение рассматривалось при дополнительном условии оптимальности по критерию наибольшего прироста биомассы фитопланктона в течение фиксированного времени.

Результаты: Учитывая изменчивость окружающей среды, был проведен ряд вычислительных экспериментов. Условия изменения освещенности и периоды отсутствия света моделировались при прочих постоянных условиях. Поведение квоты хлорофилла и доли АТФ в биомассе, рассчитанное по модели, находится в полном соответствии с представленными результатами лабораторных экспериментов [4]. Ввод в модельную систему уравнения динамики АТФ позволяет избежать прямого влияния квоты хлорофилла на внутриклеточное содержание биогенов. Таким образом формируется временной лаг, за который система успевает накопить энергию, необходимую для запуска ферментативной реакции. Характерно, что хлорофилл начинает расходоваться только тогда, когда накоплено достаточное количество энергии. В отсутствие освещения клетки фитопланктона в качестве источника энергии накапливают только АТФ. Реакция синтеза органического вещества протекает медленнее. Поэтому хлорофилл тоже расходуется медленно. Следовательно, наибольшая его концентрация в клетках фитопланктона фиксируется в абсолютной темноте [5]. Концентрация АТФ, наоборот, падает до минимума, так как в темноте потребление энергоемких веществ увеличивается из-за отсутствия других источников энергии [6].

Для решения задачи оптимального управления сформулирована и решена соответствующая краевая задача. Полученное решение достаточно близко повторяет решение исходной задачи Коши с характерным весенним и осенним всплеском биомассы. При этом максимальные значения сезонных концентраций либо сопоставимы по абсолютным значениям (весенняя вспышка), либо равны (осенняя вспышка) (рис. 1).

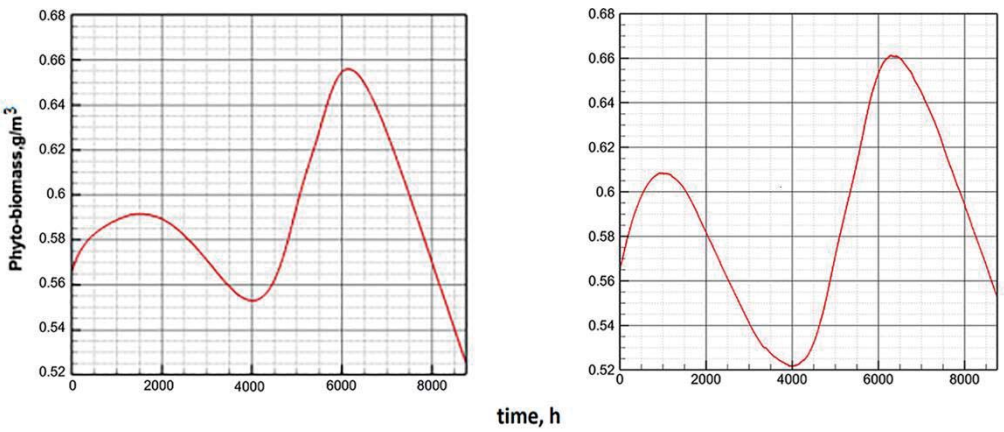


Рис. 1. Годовая динамика биомассы фитопланктона, рассчитанная по исходной модели (слева) и по модели оптимального управления (справа)

Таким образом, при условии принадлежности начально-краевых условий аттрактору модельной системы численное решение задачи оптимального управления можно использовать для прогнозирования динамики фитопланктона. При этом размерность пространства компонентов системы уменьшается. Это может существенно упростить сбор экспериментальных данных при переводе моделей в область практического использования. В то же время не следует игнорировать тот факт, что численное решение краевых задач связано с проблемами сходимости вычислительных процедур. Это требует дополнительных алгоритмов регуляризации, которые не всегда могут быть эффективно применены.

Выводы: Модели имеют сложную динамику, когда устойчивых равновесий не существует. Эта динамика подчиняется аттрактору модели Друпа, который представляет собой некоторое множество. Изучение свойств моделей, в частности, исследование устойчивости ненулевых равновесных решений, а также возможность сведения одной из них к модели оптимального управления дают основание предполагать использование модельных систем для описания реальных биологических объектов. Моделирование годового цикла показывает адекватную динамику в обеих моделях с учетом сезонных изменений среды обитания в течение года. Показано, что такие изменения приводят к весенним и осенним всплескам производства биомассы. Проведенные численные эксперименты согласуются с опубликованными результатами полевых наблюдений [7]. Учитывая наличие данных дистанционных наблюдений и результатов лабораторных экспериментов, представленные модели могут быть использованы для оценки биологической продуктивности водных экосистем.
Финансирование: Работа выполнена в рамках государственного задания ИАПУ ДВО РАН (тема № FWFW-2021-0004).

The role of photosynthetic processes in the life of phytoplankton. Mathematical model

Abakumov A., Pak S.*
Institute of Automation and Control Processes, FEG RAS, Vladivostok, Russia
** packsa@iacp.dvo.ru*

Key words: mathematical model; differential equations; phytoplankton; photosynthesis; chlorophyll; photosynthetic activity radiation; temperature; optimal control problem

Motivation and Aim: Phytoplankton is a key link in the global carbon cycle. The study of the mechanisms of primary production formation is paramount. Methods to calculate primary production using the assimilation number are known [1, 2]. This scheme is simplified because it does not take into account the variability of photo-synthetic substances in phytoplankton cells. The main one among them is chlorophyll. The ratio of chlorophyll to biomass varies depending on the state of cells and environ-mental conditions The ratio of chlorophyll to phytoplankton is usually called the chlorophyll quota [3]. All biochemical processes in biomass occur with the participation of energy-intensive substances, macroergs. They are present in living cells primarily in the adenosine triphosphate (ATP) form of. Their concentration can be measured directly today. They can be used to evaluate the characteristics of production processes. The previous stage of research consisted of developing mathematical models in which phytoplankton production is calculated taking into account the variability of the chlorophyll quota, as well as the concentration of energy-intensive substances. The study of model properties is aimed at optimizing solution methods in order to expand the scope of application of the model to real biological systems and simplify the data collection process.
Methods and Algorithms: We based our constructions on the Droop model. Traditionally, the functioning of phytoplankton in the aquatic environment is described by the Monod dependence. In same times, microorganism growth continues for some time after the nutrient resources are depleted. Monod's kinetics are unable to imitate this phenomenon. Droop's concept makes it possible to separate the nutrient uptake rate from the

growth rate. Biomass growth during the dark stage of photosynthesis is described in a similar way. The Lyapunov stability of model system equilibrium solutions is studied. Phase portraits in the spaces of key model parameters have been studied. Studying the behavior of a model on an attractor gives a more complete understanding of its properties. The solution was considered under the additional condition of optimality according to the criterion of the greatest increase in phytoplankton biomass over a fixed time.

Results: A number of computational experiments were carried out taking into account environmental variability. Conditions of changing illumination and periods of absence of light were simulated under other constant conditions. The numerical experiment demonstrates the behavior of the model components corresponding to the description of the results of a laboratory experiment under similar conditions [4]. Introducing the ATP dynamics equation into the model system allows us to avoid the direct influence of the chlorophyll quota on the intracellular content of nutrients. This creates a time lag during which the system manages to accumulate the energy necessary to start the enzymatic reaction. Chlorophyll begins to be consumed only when a sufficient amount of energy is accumulated. The reaction of organic matter synthesis is slower. Therefore, chlorophyll is also consumed slowly. Consequently, its highest concentration in phytoplankton cells is recorded in absolute darkness [5]. The ATP concentration, on the contrary, drops to a minimum, since the consumption of energy-intensive substances increases in the dark due to the absence of other energy sources [6]. To solve the optimal control problem, a corresponding boundary value problem is formulated and solved. The obtained solution of the boundary value problem quite closely repeats the solution of the initial Cauchy problem for models C and E with a characteristic spring and autumn burst of biomass. Moreover, the maximum values of seasonal concentrations are either comparable in absolute values (spring outbreak) or equal (autumn outbreak) (Fig. 1).

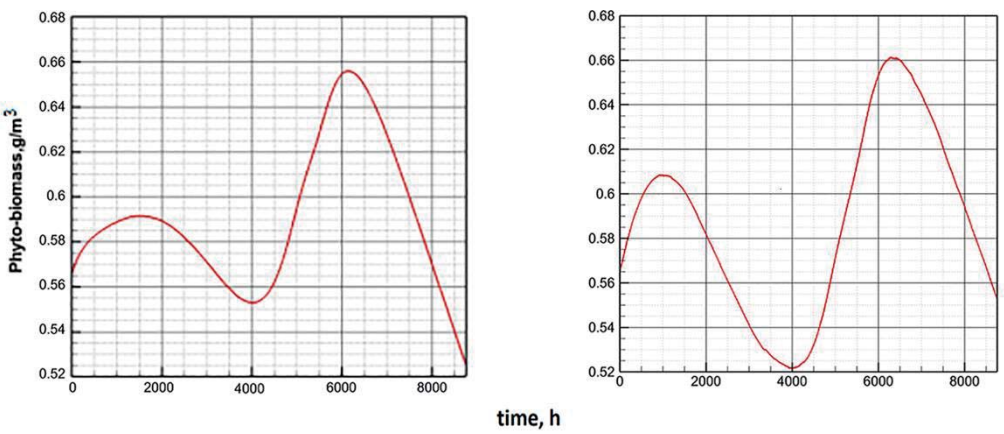


Fig. 1. Annual dynamics of phytoplankton biomass in terms of the original model (left) and according to the optimal control model (right)

Thus, under the condition that the initial-boundary conditions belong to the attractor of the model system, the numerical solution of the optimal control problem can be used to predict phytoplankton dynamics. In this case, the dimension of the space of the system components decreases. At the same time, one should not ignore the fact that the numerical solution of boundary value problems is associated with problems of convergence of computational procedures. This requires additional regularization algorithms, which cannot always be effectively applied.

Conclusion: The models have complex dynamics when the stabile equilibria don't exist. These dynamics are submitted to the attractor of the Droop model. The study of the properties of models, in particular, the study of the stability of positive equilibrium solutions, as well as the possibility of reducing one of them to an optimal control model, gives grounds to assume the use of model systems to describe real biological objects. Modeling of the annual cycle shows adequate dynamics in both models. Seasonal habitat changes lead to spring and autumn bursts of biomass productivity. This agrees with known field observations [7]. Given the availability of data from remote observations and the results of laboratory experiments, the presented models can be used to assess the biological productivity of aquatic ecosystems.

Funding: The research was carried out within the state assignment of IACP FEB RAS (Theme FWFW-2021-0004).

Список литературы/References

1. Finenko Z.Z., Suslin V.V., Churilova T.Y. The regional model to calculate the Black Sea primary production using satellite color scanner SeaWiFS. *Mar Ecol J.* 2009;8(1):81-106
2. Glover H.E. Assimilation numbers in cultures of marine phytoplankton. *J. Plankton Res.* 1980;2(1):69-79
3. Nikolaou A., Hartmann P., Sciandra A., Chachua B., Bernard O. Dynamic coupling of photoacclimation and photoinhibition in a model of microalgae growth. *J Theor Biol.* 2016;390:61-72. doi 10.1016/j.jtbi.2015.11.004
4. Hunter B.L., Laws E.A. ATP and chlorophyll a as estimators of phytoplankton carbon biomass. *Limnol Oceanogr.* 1981;26(5):944-956
5. Adamson H.Y., Hiller R.G., Vesk M. Chloroplast development and the synthesis of chlorophyll a and b and chlorophyll protein complexes I and II in the dark in Tradescantia albiflora (Kunth). *Planta*. 1980;150(4):269-274
6. Post A.F., Loogman J.G., Mur L.R. Regulation of growth and photosynthesis by Oscillatoria agardhii grown with a light/dark cycle. *FEMS Microbiol. Ecol*. 1985;1(2):97-102
7. Martinez E., Antoine D., d'Ortenzio F., de Boyer Montégut C. Phytoplankton spring and fall blooms in the North Atlantic in the 1980s and 2000s. *J Geophys Res: Oceans*. 2011;116(C11). doi 10.1029/2010JC006836

doi 10.18699/bgrs2024-2.2-02

Динамика популяции с последовательными изменениями генотипов

Братусь А.*, Непогодин А., Самокатов В.
Российский университет транспорта, Москва, Россия
** alexander.bratus@yandex.ru*

Ключевые слова: последовательные мутации бесконечно большого числа генотипов; краевая задача для уравнения потока и квазилинейного параболического уравнения

Мотивация и цель: Динамика популяций многих биологических видов характеризуется последовательными мутациями, в результате которых возникают новые генотипы видов. Типичными примерами такого процесса являются динамика популяции ВИЧ инфекции и эпидемии Ковид 19 [1, 2]. Аналогичные процессы наблюдаются в динамике популяций раковых клеток [3, 4]. Так, например, в результате последовательных мутаций возникают 64 раковые и предраковые клетки поджелудочной железы [5]. Представленная в работе математическая модель динамики популяции с последовательными мутациями основывается на следующих предположениях.

1. Множество генотипов является бесконечно большим, и каждому генотипу присваивается номер x из множества $[0, +\infty)$.
2. Показатели роста генотипа зависят от времени.
3. Динамика плотности численности генотипов описывается логистическим уравнением, в котором предельная численность популяции зависит от вида (номера) генотипа.
4. Последовательные мутации и появление новых генотипов происходят с заданной скоростью $v(t) \geq 0, t \geq 0$.
5. Наряду с последовательными мутациями возникают обратимые мутации малой интенсивности [6].

Предложенная модель не учитывает случайный характер постоянно возникающих мутаций. Однако большинство этих мутаций носят тупиковый характер и в процессе эволюции реализуются лишь те мутации, которые способствуют успешному выживанию видов.

Результаты: В работе показано, что при сделанных предположениях динамика популяции с последовательными изменениями генотипов описывается краевой задачей для уравнения потока, которая во многих случаях допускает точные аналитические решения.

Предложенная модель допускает обобщение на случай, когда имеется несколько независимых ветвей последовательных мутаций генотипов.

Dynamics of population with consecutive mutation of genotypes

Bratus A.*, Nepogodin A., Samokatov V.
Russian University onTtransport, Moscow, Russia
** alexander.bratus@yandex.ru*

Key words: consecutive mutations of infinitely many genotypes; boundary value problem for continuity equation and quasilinear parabolic equation

Motivation and Aim: Population dynamics of many biological species is characterized by consecutive mutations that result in emergence of new genotypes. Typical examples of such process include population dynamics of HIV infections and Covid-19 epidemic [1, 2]. Similar processes are observed in population dynamics of cancer cells [3, 4]. For example, as a result of consecutive mutations of pancreatic cells 64 types of cancer and pre-cancer cells emerge [5]. In this work the mathematical model of population dynamics with consecutive mutations is based on the following assumptions.

1. The set of genotypes is infinitely large, and each genotype is given a number x from $[0, +\infty)$.
2. The growth rate of genotypes depends on time.
3. The dynamic of density of the number of genotypes is described by a logistic equation where the population limit depends on the number of a genotype.
4. Consecutive mutations and emergence of the new genotypes happen with a given speed $v(t) \geq 0, t \geq 0$.
5. In addition to the consecutive mutations a reverse mutations of low intensity appear [6].

The proposed model does not account for the randomness of constantly appearing mutations. Nevertheless, the majority of such mutations are dead ends, and in the process of evolution only mutations with facilitate the successful survival of the species actualize.

Results: In this work it is shown that following these assumptions the population dynamics with consecutive mutations of genotypes is described with boundary value problem for continuity equation, which in many cases has analytical solutions.

The proposed model allows generalizations for describing cases with multiple independent branches of consecutive mutations.

Список литературы/References

1. Martinez J.P., Bocharov G., Ignatovich A., Reitez J., Dittmar M. Fitness ranking of individual drives of epistatic interactions in HIV-1. *PloS One.* 2011;6(3):e18375. doi 10.1371/journal.pone.0018375
2. Blyuss K.B., Kyrychko Y.N. Mathematical model of replication – mutation dynamics in coronaviruses. *bioRxiv*. doi 10.1101/2024.01.29.577716
3. Loeb L.A. Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat Rev Cancer*. 2011;11:450-457
4. Loeb K.R., Loeb L.A. Significance of multiple mutation on cancer. *Carcinogenesis*. 2000;21(3):379-385
5. Bratus A.S., Leslie N., Chamo M., Grebennikov D., Savinkov R., Bocharov G., Yurchenko D. Mathematical model of pancreatic cancer cell dynamics considering the set of sequential mutations. *Mathematics.* 2022;10(19):3557. doi 10.3390/math10193557
6. Bessonov N., Bocharov G., Leon C., Popov V., Volpert V. Genotype-dependent virus distribution and competition of virus strains. *Math Mech Complex Systems.* 2020;8(2):101-126

doi 10.18699/bgrs2024-2.2-03

Моделирование процессов формирования видового и генетического разнообразия в замкнутых экосистемах

Букин Ю.С.[1, 2]*, Бережной М.Д.[2], Щербаков Д.Ю.[1, 3]

[1] *Лимнологический институт СО РАН, Иркутск, Россия*

[2] *Иркутский государственный университет, Иркутск, Россия*

[3] *Новосибирский государственный университет, Новосибирск, Россия*

**bukinyira@mail.ru*

Ключевые слова: эволюция; видообразование; математическое моделирование; генотип; генетическое разнообразие; случайные процессы

Введение: В составе биосферы Земли существует такое понятие, как изолированная экосистема. Изолированные экосистемы активно обменивались с окружающей средой веществом и энергией и в гораздо меньшей степени обменивались с другими экосистемами организмами. Если эта экосистема существует в течение длительного периода времени, то в результате биологической эволюции в ней накапливается большое количество эндемичных видов. Часто образуются букеты, содержащие несколько эндемичных видов или даже родов и семейств. Примерами таких экосистем являются древние озера (Байкал, Танганьика, Каспийское море); глубокие выходы и грязевые вулканы; островные экосистемы.

Процессы видообразования в таких экосистемах происходят как на микроуровне (изменение частот встречаемости генотипов), так и на макроуровне (формирование букетов эндемичных видов, родов и семейств). Микроэволюционный уровень заключается в появлении новых генотипов благодаря мутационному процессу и фиксации новых генотипов в популяции, если они обладают эволюционным преимуществом. Таким образом может происходить адаптивное преобразование существующих видов и эволюционная дивергенция – появление новых видов. При макроэволюции в замкнутой экосистеме происходит перераспределение потенциально большого количества свободных экологических ниш за достаточно короткие в геологическом масштабе отрезки времени между большим количеством новых видов, появляющихся в ходе макроэволюции.

Практически все типы эволюционных процессов можно изучить с помощью методов имитационного компьютерного моделирования. В нашем исследовании мы разработали несколько имитационных моделей, описывающих как процессы микроэволюции, так и процессы макроэволюции в замкнутых экосистемах.

Методы и алгоритмы: В исследовании применялся подход индивидуально-ориентированного (объектно-ориентированного) моделирования разработки имитационных компьютерных моделей. Для наших разработок мы взяли за основу подход индивидуально-ориентированного моделирования из работ [1–3].

Результаты: Нами были разработаны несколько моделей микроэволюционных процессов. Модель симпатрического видообразования (видообразование без пространственного разделения организмов) с внедрением нейтральных генетических маркеров для отслеживания истории видообразования. Данная модель позволила отследить ранние этапы формирования дивергенции в нейтральных маркерах при параллельном накоплении адаптивных преобразований в ненейтральных локусах

[4]. Модель парапатрического видообразования, включающая пространственное разобщение особей в условиях однородного и неоднородного ареала обитания. Данная модель позволила изучить вклад пространственного разобщения в процесс конкурентного видообразования и эволюцию нейтральных генетических маркеров [5–7]. Модель эволюции двух видов, взаимодействующих по типу хищник – жертва (паразит – хозяин). Модель эволюции хищник–жертва позволила сделать вывод, что воздействие хищника замедляет процессы конкурентного видообразования у жертвы [8].

Для исследования процессов макроэволюции нами была разработана имитационная модель, описывающая формирование букета видов. Изначально в модели предполагалось существование одного вида и определенного количества свободных экологических ниш. С некоторой вероятностью с течением времени вид может эволюционно дивергировать и занять новую нишу, переселив в нее дочерний вид. Далее процесс развивается по ветвящейся цепочке. Модель позволила изучить закономерности заселения всей экосистемы (всех свободных ниш) в зависимости от вероятности эволюционной дивергенции и количества экологических ниш.

Финансирование: Работа поддержана темой бюджетного финансирования ЛИН СО РАН «Генетика сообществ байкальских организмов: структура генофонда, стратегии консервации» 0279-2021-0010 (121032300196-8).

Speciation process and genetic diversity formation modeling in closed ecosystems

Bukin Yu.S.[1, 2]*, Berezhnoi M.D.[2], Sherbakov D.Yu.[1, 3]

[1] *Limnological Institute, SB RAS, Irkutsk, Russia*

[2] *Irkutsk State University, Irkutsk, Russia*

[3] *Novosibirsk State University. Novosibirsk, Russia*

* *bukinyira@mail.ru*

Key words: evolution; speciation; mathematical modelling; genotype; genetic diversity; random processes

Motivation and Aim: Within the Earth's biosphere there is such a thing as an isolated ecosystem. Isolated ecosystems actively exchanged matter and energy with the environment, and to a much lesser extent exchanged organisms with other ecosystems. If this ecosystem exists for a long period of time, then as a result of biological evolution, a large number of endemic species accumulate in it. Often bouquets are formed containing several endemic species or even genera and families. Examples of such ecosystems are ancient lakes (Lake Baikal, Tanganyika, Caspian Sea); deep outcrops and mud volcanoes; island ecosystems.

The processes of speciation in such ecosystems occur both at the micro level (changes in the frequency of occurrence of genotypes) and at the macro level (formation of bouquets of endemic species, genera and families). The microevolutionary level consists of the emergence of new genotypes due to the mutation process and the fixation of new genotypes in the population if they have an evolutionary advantage. In this way, adaptive transformation of existing species and evolutionary divergence – the emergence of new species – can occur. During macroevolution in a closed ecosystem, a potentially large number of free ecological niches are redistributed over fairly short periods of time on a

geological scale between a large number of new species appearing during macroevolution.

Almost all types of evolutionary processes can be studied using computer simulation methods. In our research, we developed several simulation models that describe both microevolutionary and macroevolutionary processes in closed ecosystems.

Methods and Algorithms: The study used an individual-based (object-oriented) modeling approach to the development of simulation computer models. For our developments, we took as a basis the approach of individual-based modeling from the works [1–3].

Results: We have developed several models of microevolutionary processes. Model of sympatric speciation (speciation without spatial separation of organisms) using neutral genetic markers to trace the history of speciation. This model made it possible to track the early stages of the formation of divergence in neutral markers with the parallel accumulation of adaptive transformations in non-neutral loci [4]. The model of parapatric speciation, including spatial separation of individuals in homogeneous and non-uniform habitats. This model made it possible to study the contribution of spatial separation to the process of competitive speciation and the evolution of neutral genetic markers [5–7]. The model of the evolution of two species interacting as a predator – prey (parasite – host). The model of predator-prey evolution led to the conclusion that the influence of a predator slows down the processes of competitive speciation in the prey [8].

To study the processes of macroevolution, we developed a simulation model that describes the formation of a bouquet of species. Initially, the model assumed the existence of one species and a certain number of free ecological niches. With some probability, over time, a species may diverge evolutionarily and occupy a new niche, moving a daughter species into it. The process then develops along a branching chain. The model made it possible to study the patterns of settlement of the entire ecosystem (all free niches) depending on the probability of evolutionary divergence and the number of ecological niches.

Funding: Work supported by budget funding of Limnological Institute SB RAS “Genetics of communities of Baikal organisms: structure of the gene pool, conservation strategies” 0279-2021-0010 (121032300196-8).

Список литературы/References

1. Dieckmann U., Doebeli M. On the origin of species by sympatric speciation. *Nature*. 1999;400(6742):354-357. doi 10.1038/22521
2. Doebeli M., Dieckmann U. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *Am Nat*. 2000;156(S4):S77-S101. doi 10.1086/303417
3. Takasu F. Co-evolutionary dynamics of egg appearance in avian brood parasitism. *Evol Ecol Res.* 2003;5(3):345-362
4. Semovski S.V., Verheyen E., Sherbakov D.Yu. Simulating the evolution of neutrally evolving sequences in a population under environmental changes. *Ecol. Modell.* 2004;176(1-2):99-107. doi 10.1016/j.ecolmodel.2003.07.013
5. Semovski S.V., Bukin Yu.S., Sherbakov D.Yu. Speciation and neutral molecular evolution in one-dimensional closed population. *Int J Mod Phys C*. 2003;14(07):973-983. doi 10.1142/S012918310300511X
6. Bukin Yu.S., Pudovkina T.A., Sherbakov D.Yu., Sitnikova T.Ya. Genetic flows in a structured one-dimensional population: simulation and real data on Baikalian polychaetes M. godlewskii. *In Silico Biol.* 2007;7(3):277-284. doi 10.1111/j.1365-294X.2009.04092.x
7. Bukin Yu.S., Gorbylev A.L. An Individual-Based Model to Simulate Genetic Processes in Populations of Species Inhabiting One-Dimensional Area. *Matematicheskaya Biologiya i Bioinformatika*. 2014;9(2):438-452. doi 10.17537/2014.9.438
8. Bukin Yu S. Coevolution in a predator-prey system: An ecogenetic model. *Russ J Genet: Appl Res.* 2014;4:543-548. doi 10.1134/S2079059714060045

doi 10.18699/bgrs2024-2.2-05

Микробное сообщество: определяющие свойства и закономерности в динамике и стационаре

Дегерменджи А.[1], Абакумов А.[2]*

[1] *Институт биофизики ФИЦ "Красноярский научный центр СО РАН", Красноярск, Россия*

[2] *Институт автоматики и процессов управления ДВО РАН, Владивосток, Россия*

* *abakumov@dvo.ru*

Ключевые слова: хемостат; моделирование; управление структурой сообщества; эффект аутостабилизации; квантование коэффициентов

Мотивация и цель: В работе обобщены теоретические и экспериментальные исследования авторов, направленные на изучение закономерностей устойчивого сосуществования взаимодействующих микробных популяций в пределах одного трофического уровня. Популяции взаимодействуют в открытых непрерывных культурах посредством регулирующих факторов.

Методы и алгоритмы: В математических моделях микробных сообществ [1, 2] исследуются свойства динамики видового состава, изучаются закономерности и особенности устойчивого сосуществования популяций, взаимодействующих посредством регулирующих факторов (RF). Под регулирующими факторами понимаются субстраты и менее обычные биологические и физико-химические характеристики: метаболиты, ингибирующие рост; уровень pH среды; температура среды обитания и т. д.

Результаты: Анализ математических моделей смешанных микробных потоковых культур, взаимодействующих через RF, устанавливает правило: в равновесии число сосуществующих популяций не превышает числа RF, уровень входа которых контролирует состав сообщества. Регулирующие факторы демонстрируют эффект автостабилизации: их равновесные уровни в системе не связаны с их входными уровнями. Этот эффект обнаруживается через коэффициенты чувствительности. Закон для всего сообщества в идеальной однородной потоковой системе гласит, что сумма всех коэффициентов чувствительности представляет собой целое число, равное разности между числом RF и количеством сосуществующих популяций. В природных экосистемах с их изменяющимися условиями эффект автостабилизации проявляется в виде тенденции к отсутствию корреляции между фоновой концентрацией RF и численностью контролируемой популяции из-за обычного отсутствия данных о входных уровнях RF. Теоретические условия сосуществования видов достаточно сложны, а их реализация в природе сомнительна: в работах [3, 4] «хаотическое» состояние динамики популяций имитируется сложной комбинацией дискретных и случайных процессов. Для моделей, явно учитывающих RF, в самом общем случае для произвольных функций и типов взаимодействий строго доказан «принцип исключения»: даже динамическое сосуществование двух видов с одним RF любой природы невозможно [5].

Выводы: Эффект автостабилизации – это редкий пример теоретически точно выведенного инварианта или «закона экологии» для всей экосистемы, который был получен для сообщества с обширной и сложной сетью «зависимых от плотности»

взаимодействий. Используя на практике формулы квантования, мы можем оценить полноту наших знаний о системе взаимодействия популяций в сообществе в рамках достаточно общих предположений о «росте» и возрастной структуре каждой популяции. Результаты экспериментов позволяют непосредственно стремиться к поиску неучтенных типов питательных веществ или количества микроорганизмов по степени приближения к выполнению условия квантования. Регулирующие факторы демонстрируют эффект автостабилизации: их равновесные уровни в системе не связаны с их входными уровнями. В природных экосистемах с их изменяющимися условиями эффект автостабилизации проявляется в виде тенденции к отсутствию корреляции между фоновой концентрацией РФ и численностью контролируемой популяции.
Финансирование: Работа поддержана грантом РНФ, № 23-44-00059, https://rscf.ru/en/project/23-44-00059/.

Microbial community: defining properties and patterns in dynamics and steady state

Degermendzhi A.[1], Abakumov A.[2]*

[1] *Institute of Biophysics, Department of Federal Research Center "Krasnoyarsk Science Center of the Siberian Branch of the Russian Academy of Sciences", Krasnoyarsk, Russia*

[2] *Institute for Automation and Control Processes, FEB RAS, Vladivostok, Russia*

* *abakumov@dvo.ru*

Key words: chemostat; modelling; control of community composition; autostabilization effect; coexistence; new interaction criterion

Motivation and Aim: The paper summarizes the authors' theoretical and experimental studies aimed at studying the patterns of sustainable coexistence of interacting microbial populations within the same trophic level. Populations interact in open continuous cultures through regulating factors.

Methods and Algorithms: The properties of the dynamics of species composition in mathematical models of microbial communities, and the patterns of sustainable coexistence of populations interacting through regulatory factors (RF) are studied [1, 2]. The regulatory factors are substrates and biological, physic, and chemical characteristics: growth-inhibiting metabolites; pH level of the environment; habitat temperature, etc.

Results: Analysis of mathematical models of mixed microbial flow cultures interacting through RFs establishes the rule: the number of coexisting populations in equilibrium does not exceed the number of RFs, the level of entry of which controls the composition of the community. Regulating factors exhibit an unexpected autostabilization effect: their equilibrium levels in the system are not related to their input levels. This effect is detected through sensitivity coefficients. The law for the entire community in an ideal homogeneous flow system states that the sum of all sensitivity coefficients is an integer. It is the difference between the number of RFs and the number of coexisting populations. In natural ecosystems with their changing conditions, the autostabilization effect will manifest itself as a tendency for there to be no correlation between background RF concentration and monitored population size. The theoretical conditions for the coexistence of species are quite complex, and their implementation in nature is questionable: in the

works [3, 4] “chaotic” coexistence of population dynamics is imitated by a complex combination of discrete and random processes. For models that explicitly consider RF, in the most general case, for arbitrary functions and types of interactions is strictly proven: even the dynamic coexistence of two species with one RF of any nature is impossible [5].

Conclusion: The autostabilization effect is a example of a theoretically rigorously derived invariant or “law of ecology” for an entire ecosystem that has been derived for a community with a vast and complex network of “density-dependent” interactions. Using quantization formulas in practice, we can assess the completeness of our knowledge about the system of populations interaction in a community within the framework of fairly general assumptions about the “growth” and age structure of each population. The results of the experiments allow us to directly strive to search for unaccounted types of nutrients or the number of microorganisms according to the degree of approximation to the fulfillment of the quantization condition. Regulating factors exhibit an unexpected autostabilization effect: their equilibrium levels in the system are not related to their input levels. In natural ecosystems with their changing conditions, the autostabilization effect will manifest itself in the form of a tendency towards a lack of correlation between the background RF concentration and the size of the controlled population.

Funding: The study was supported by Grant of the Russian Science Foundation No. 23-44-00059, https://rscf.ru/en/project/23-44-00059/.

Список литературы/References

1. MacArthur R. Species packing and competitive equilibrium for many species. *Theor Pop Biol*. 1970;1:1-11
2. Armstrong R.A., McGehee R. Coexistence of species competing for shared resources. *Theor Pop Biol*. 1976;9:317-328
3. Hutchinson G.E. The paradox of the plankton. *Am Nat.* 1961;95:137-145
4. Degermendzhi A.G. On the problem of coexistence of interacting continuous populations. In: Pechurkin N. (Ed.). Mixed continuous cultures of microorganisms. Novosibirsk: Nauka, 1981;26-106 (in Russian)
5. Degermendzhi A.G., Abakumov A.I. The Principle of Competitive Exception in a Two-Species Community with One Metabolic Regulation Factor. *Doklady Biochemistry and Biophysics.* 2018;480(1):149-151. doi 10.1134/S1607672918030055

doi 10.18699/bgrs2024-2.2-06

Разработка методов и средств анализа биологического разнообразия в бассейне Черного и Азовского морей на основе интеграции междисциплинарной информации

Кулешова О.*, Кривенко О.
Институт биологии южных морей им. А.О. Ковалевского РАН, Севастополь, Россия
** v_olgo4ka@inbox.ru*

Ключевые слова: биологическое разнообразие; распределенные системы; интеграция данных

Мотивация и цель: Интеграция разноплановой информации о гидробионтах и местах их обитания является первичным этапом создания эффективного инструмента изучения биологического разнообразия, необходимого как для понимания фундаментальных вопросов эволюции жизни на Земле, так и для решения прикладных задач в области сохранения морских экосистем [1]. Попытки такой интеграции на глобальном уровне предпринимаются в рамках мировых баз данных генетической (NCBI) и таксономической (WoRMs, ORBIS) информации [2]. Но для решения практических задач моделирования структуры биоразнообразия и связанного с ней функционирования морских экосистем в условиях изменяющегося климата и антропогенной нагрузки требуется обобщение разнообразной по содержанию и форме информации, имеющей достаточное пространственное и временное разрешение. Получить полный спектр таких данных возможно только на уровне отдельных регионов, дополнительно обобщая и анализируя локальные источники научной информации. Особенно актуально изучение акваторий, устойчивое функционирование которых находится под угрозой из-за негативного воздействия климатических и антропогенных факторов, а также экосистемы с уникальными условиями для жизни.

Черное море представляет интерес с обеих точек зрения. Во-первых, это самый большой меромиктический водоем, где аэробная зона жизни составляет лишь 10–15 % от его общего объема. За последние 40 000 лет своей геологической истории этот закрытый прибрежный бассейн претерпел несколько преобразований из пресноводного озера в морской водоем и обратно. В настоящее время соленость Черного моря приблизительно в два раза ниже, чем в океане, поэтому здесь могут выживать только эвригалинные организмы, а некоторые крупные таксоны полностью отсутствуют. Низкое видовое разнообразие в сочетании с высоким уровнем разнообразия мест обитания создает благоприятные условия для появления инвазивных видов в регионе. Эти виды могут нарушать баланс и функционирование экосистем, создавая значительную угрозу его биоразнообразию. Кроме того, коммерческий промысел и чрезмерная эксплуатация ресурсов, эвтрофикация, химическое загрязнение могут значительно усиливать негативное воздействие. Классический пример – коллапс пелагической экосистемы Черного моря в конце двадцатого века. Учитывая вышесказанное, разработка методов и средств анализа биологического разнообразия на основе интеграции междисциплинарной информации

является актуальной как для решения прикладных вопросов сохранения и рационального использования ресурсов Азово-Черноморского бассейна, так и для решения фундаментальных научных проблем.

За полтора века современной истории изучения биоразнообразия Черного моря накоплен огромный материал по таксономии, биологии, экологии гидробионтов, особенностях их физиологии, а также о структуре и функционировании отдельных сообществ и экосистемы бассейна в целом. Есть результаты молекулярно-генетических исследований и сведения о генетическом разнообразии популяций некоторых видов черноморских организмов. Часть этой информации, как исторической, так и более современной, собрана в разнообразных базах данных, но лишь небольшая их доля доступна через глобальные системы. Для реализации современных подходов к анализу биоразнообразия требуется интеграция этой разрозненной информации в рамках единой системы.

Методы и алгоритмы: Разработка системы включает несколько этапов: определение целей и постановка задач, формальное описание (спецификация), логическое проектирование и реализация. Особое внимание при проектировании должно быть уделено спецификации, так как сложность устранения ошибок, допущенных на этом этапе, многократно возрастает при их обнаружении на этапе эксплуатации. Поэтому необходим особый подход к формальному описанию системы, который обеспечит оптимальный способ получения данных, быстрый доступ к ним в дальнейшем, а также легкую совместимость данных. Кроме того, потребуется независимое формирование и возможность функционирования модулей, отвечающих за доступ к различным базам данных и сервисам, а также обеспечение возможности простого подключения дополнительных программных средств для анализа, моделирования и прогнозирования в рамках изучения биологического разнообразия и решения задач по сохранению и рациональному использованию морских ресурсов.

Для реализации мы предлагаем использовать методы и средства, применяемые для спецификаций распределенных систем. Распределенные системы – это дискретные информационные и управляющие системы, которые представляют собой сложные комплексы, состоящие из множества элементов (технических и программных средств), разнесенных в пространстве, не зависящих друг от друга, но взаимодействующих для выполнения общих задач. К их основным характеристикам относятся функциональная целостность, параллельность работы элементов, переменность структуры и другие параметры, касающиеся эффективности работы системы. Для описания динамики функционирования объектов таких систем имеется ряд моделей, включая сети Петри, асинхронные процессы, регулярные выражения, конечные автоматы. Оптимальная модель должна обеспечивать баланс между уровнем абстракции и уровнями понятности, адекватности, описательной мощности, что существенно облегчает процессы спецификации и реализации систем.

Наиболее приемлемой моделью спецификации интегративной системы для изучения биоразнообразия является инициальный детерминированный конечный автомат Рабина–Скотта на формальном уровне [3]. Через использование языка таблиц событий [4] реализуется прозрачность перехода от неформального к формальному описанию. Такой подход позволяет обеспечить понятность, полноту, непротиворечивость, модифицируемость, верифицируемость, автоматизируемость процесса проектирования и дальнейшей реализации спецификации системы.

В качестве источников таксономической, генетической и другой сопутствующей наблюдениям информации мы рассматриваем такие системы, как WoRMS (marinespecies.org), GenBank (ncbi.nlm.nih.gov), GBIF (gbif.org), OBIS (obis.org), имеющие различные утилиты доступа к наборам данных. Программные средства реализации включают в себя языки программирования в соответствии с инструкциями к утилитам доступа, такие как Perl, Python, R; API функции для доступа к данным; СУБД PostgreSQL для хранения и доступа к объединенным данным.
Результаты: Проанализированы возможности получения и интеграции данных из систем WoRMS, GenBank, GBIF, OBIS. Выявлены следующие особенности, влияющие на процессы интеграции информации. WoRMS имеет географическую привязку для таксонов, их описание и ссылки на другие ресурсы, возможность формирования набора данных с географической локализацией, но не имеет списка данных конкретных образцов. В GenBank хранятся генетические данные, описание которых может содержать указание на географический регион и(или) географические координаты, при этом отсутствует возможность формирования точной выборки по этим параметрам. В GBIF имеются средства выборки по географическому региону (водоему), однако сформированный при этом набор данных имеет лишние данные (сухопутные, не относящиеся к запрошенному водоему) и не включает информацию о генетических данных. В OBIS для Черного моря присутствуют наборы данных с широким временным диапазоном сбора проб, без привязки к генетическим данным. С учетом перечисленных особенностей мы разработали первичную спецификацию методов и средств для анализа биологического разнообразия в бассейне Черного и Азовского морей, реализация которой позволяет объединить разрозненные наборы данных, определить представленность таксонов, получить для анализа выборку имеющихся генетических данных, анализировать распределение интересующих таксонов во времени и пространстве. Такой подход обеспечивает стандартизацию и консолидацию данных из разных источников и помещение этих наборов в структуры хранения, обеспечивающие простой и оперативный доступ к этим данным.
Выводы: Выполнена первичная спецификация методов и средств для анализа биологического разнообразия в бассейне Черного и Азовского морей. Их реализация позволяет автоматизировать процессы получения и объединения разрозненных наборов данных о биоразнообразии. Дальнейший анализ полученных данных в первую очередь будет направлен на изучение встречаемости видов, выявление пробелов в генетических исследованиях и формирование соответствующих задач для исследований. Следующий этап предполагает разработку способов получения и интеграции биологической информации из более специализированных источников, материалов экспедиций научно-исследовательских судов, а также данных о гидрофизических и гидрохимических условиях среды. Объединение полученной информации с привязкой к пространственным координатам позволит обоснованно подойти к реализации задач моделирования динамики структуры биоразнообразия и функционирования экосистем на основе данных наблюдений.
Финансирование: Работа выполнена в рамках государственного задания ФИЦ ИнБЮМ № 124022400148-4.

The development of a tool to analyze biodiversity in the Azov-Black Sea region by merging multidisciplinary information

Kuleshova O.*, Krivenko O.
A.O. Kovalevsky Institute of Biology of the Southern Seas, RAS, Sevastopol, Russia
** v_olgo4ka@inbox.ru*

Key words: biodiversity; distributed systems; data integration

Motivation and Aim: The flexible integration of disparate data about aquatic organisms and their habitats is the first step toward creating a valuable tool for the comprehensive analysis of biological diversity [1]. This instrument is essential for understanding the fundamental aspects of life evolution on Earth and for solving applied challenges related to the conservation and sustainable functioning of marine ecosystems. Attempts at such integration are being made on a global scale through the use of world databases for genetic (NCBI) and taxonomic information (WoRMs, ORBIS) [2]. However, in order to address the practical challenges of modelling the structure of biodiversity and the functioning of marine ecosystems, it is necessary to integrate various data types with sufficient spatial and temporal resolution. Currently, only at the regional level is it feasible to obtain a comprehensive range of such data by additional summarization and analysis of local scientific information sources. The main fundamental interests in biodiversity research include water areas with unique living conditions. Scientists also focus on ecosystems that are threatened by the negative impacts of climate change and human activities.

The Black Sea region is of particular interest for research due to both of these reasons. It is the largest body of meromictic water on the planet, with an aerobic life zone only existing in the upper 10–15 % of its total volume. During the last 40,000 years of its geological history, this enclosed coastal basin has undergone several transformations, changing from a freshwater lake into an ocean-like body of water and then back again. About 7,000 years ago, modern flora and fauna began to develop after the Bosporus opened and Mediterranean water began flowing into the area. Nowadays, the salinity of the Black Sea is twice lower than that of the ocean. This allows only euryhaline organisms to survive here. Some large taxa are completely absent from the Black Sea. Low species diversity, combined with high levels of habitat diversity in the region, creates favorable conditions for the establishment of invasive species. These invasive species can disrupt the balance and functioning of ecosystems, posing a significant threat to biodiversity in the Black Sea area. In addition, commercial fishing and overexploitation of resources, as well as chemical contamination, eutrophication and pollution, can significantly increase the negative impact on biodiversity in the region. A classic example of this phenomenon is the collapse of the pelagic ecosystem in the Black Sea during the late twentieth century. Considering all of the above, the Black Sea is an urgent case study for the development of a helpful tool for the comprehensive analysis of biological diversity.

Over a century and a half of modern history, studying biodiversity in the Black Sea has gathered vast amounts of material on the taxonomy, biology, and ecology of aquatic organisms. This includes information on the peculiarities of the physiology, structure,

and function of communities and ecosystems in the basin, as well as the molecular genetics and genetic diversity of some populations of Black Sea species. Information has been collected from various databases, though only a small proportion is available through global systems. To implement modern approaches to biodiversity analysis, all this diverse information needs to be integrated into a single system.

Methods and Algorithms: The development of the system involves several main stages, including the definition of goals and objectives, formal description (specification), logical design, and implementation. Special attention should be given to the specification stage, as errors made during this stage can be difficult to fix later on. Therefore, a special approach is required for the formal description of the system in order to provide an optimal method for obtaining data and easy access to it in the future, as well as facilitate data compatibility. Additionally, it should allow for the independent operation of modules responsible for accessing various databases and services. It will also make it easier to integrate additional software tools for analysis, modelling, and forecasting into the context of studying biological diversity and its protection.

We propose to achieve this goal by using the methods and tools used in the design of distributed systems. A distributed system is a complex system that consists of many interconnected elements, both hardware and software, that work together to perform a common task. These systems are characterized by their functional integrity, parallelism, variability, and other features that contribute to their overall effectiveness. To describe the behavior of these systems, we can use a variety of models, including Petri nets, asynchronous processes, and finite state machines. The optimal model will depend on the balance between abstraction and clarity, accuracy, and descriptive power, which will facilitate the specification and implementation processes.

The most suitable model for the design of an integration system to study biodiversity is the initial, Rabin–Scott deterministic finite state machines at the formal level [3]. Through the use of the event table language [4], we can ensure the transparency of the transition from an informal to a formal description. This approach allows us to ensure clarity, completeness, consistency, modifiability, and verifiability of the design process, as well as the further implementation of the system specification. We consider such sources of taxonomic, genetic, and other observational data as WoRMS (marinespecies.org), GenBank (ncbi.nlm.nih.gov), GBIF (gbif.org), OBIS (obis.org), which have various tools for accessing datasets. For software implementation, we use programming languages according to the instructions for accessing these tools, such as Perl, Python, and R; API functions for data access, and PostgreSQL DBMS to store and access aggregated data.

Results: The possibilities of obtaining and data integration from various systems, such as WoRMS, GenBank, GBIF, and OBIS, have been analyzed. Several features have been identified that influence the process of information integration. WoRMS provides geographical references for taxa, descriptions, and links to other resources. It also has the ability to create a dataset with geographical localizations, but it does not provide a list of specific sample data. GenBank stores genetic data that may include information about a region or coordinates. However, it is not possible to create an accurate sample based on these parameters. GBIF has sampling tools for specific geographical regions, but the dataset generated from this process includes superfluous land-based data unrelated to the requested region, and it does not contain information about genetic data. OBIS for the Black Sea provides datasets with a wide range of collection dates, but without reference to genetic data. Considering these features, a primary specification of methods and tools

has been developed for the analysis of biological diversity in the Black Sea and Azov region. The implementation of these methods makes it possible to combine various data sets, determine the presence of taxa, searching genetic data, and analyze taxa distribution in time and space. This approach ensures the standardization and consolidation of data from different sources, as well as the storage of this data in structures that allow easy and quick access to it.

Conclusion: The primary specification of methods and tools for analyzing biological diversity in the Black and Azov Seas basin has been done. Their implementation allows automating the processes of obtaining, combining disparate data sets on the biodiversity. Their further analysis will primarily focus on studying the occurrence of species, identifying gaps in genetic research, and forming appropriate research objectives.

The next stage involves developing ways to obtain and integrate biological information from specialized sources, such as materials from research vessel expeditions, as well as data on environmental conditions. Combining the information obtained with spatial coordinates will allow us to reasonably approach the task of modeling the dynamics of biodiversity structure and the functioning of ecosystems.

Funding: The study is supported by IBSS GA (No. 124022400148-4).

Список литературы/References

1. Webb T.J., Vanhoorne B. Linking dimensions of data on global marine animal diversity. *Phil Trans. R Soc*. 2020;375(1814):20190445. doi 10.1098/rstb.2019.0445
2. Bouchet P., Decock W., Lonneville B., Vanhoorne B., Vandepitte L. Marine biodiversity discovery: the metrics of new species descriptions. *Front Mar Sci.* 2023;10:929989. doi 10.3389/fmars.2023.929989
3. Апраксин Ю.К. Управление информационным взаимодействием в распределенных технических системах: конечно-автоматный подход. М.: Инфра-М, 2018
   [Apraksin Yu.K. Information interaction management in distributed technical systems: a finite-automaton approach. M.: Infra-M, 2018 (in Russian)]
4. Кулешова О.Н. Разработка методов спецификации информационных моделей средствами языка таблиц событий. *Восточно-Европейский журнал передовых технологий*. 2012;4(2):28-31
   [Kuleshova O.N. Development of methods for specification of information models by means of the event table language. *Eastern European Journal of Advanced Technologies*. 2012;4(2):28-31 (in Russian)]

doi 10.18699/bgrs2024-2.2-07

Применение методов машинного обучения для анализа генетического полиморфизма популяций

Марьяновская Т.А.[1]*, Щербаков Д.Ю.[1, 2]

[1] *Новосибирский государственный университет, Новосибирск, Россия*

[2] *Лимнологический институт СО РАН, Иркутск, Россия*

* *t.maryanovskaya@g.nsu.ru*

Ключевые слова: микроэволюция; популяционная генетика; искусственные нейронные сети

Мотивация и цель: Организмы эволюционируют в составе сообществ разных видов, и для изучения микроэволюции было разработано много моделей в рамках дисциплины популяционной генетики. Применение методов глубокого обучения становится все более актуальным в биологии из-за большого объема накопленных генетических данных. Анализ микросателлитов представляет определенные трудности из-за их сложной структуры. Это обусловлено переменной длиной микросателлитных последовательностей и их повторяющейся природой. Изучение и анализ микросателлитов требуют специальных методов и подходов для интерпретации полученных данных в генетических исследованиях.

Применение глубокого обучения в генетике популяций открывает новые возможности для анализа генетических данных и для анализа микроэволюционных процессов. Целью этой работы является исследование популяционной динамики у организмов со смешанной стратегией размножения с помощью методов глубокого обучения.

Методы и алгоритмы: В данном исследовании было использовано стохастическое моделирование. Спецификация модели представляет собой сверточную нейронную сеть с 6 внутренними и 2 внешними слоями. Обучение этой модели производилось на данных, полученных от модифицированной модели Райта–Фишера, включающей дополнительное бесполое размножение.

Результаты: В данной работе была разработана модель, представляющая одну из первых попыток применения глубокого обучения для задач популяционной генетики. Модель осуществляет анализ экспериментальных данных и прогнозирует соотношение полового и бесполого размножения у организмов, способных изменять свою стратегию размножения, таких как гидры, некоторые ракообразные, моллюски и другие. Проведены вычислительные эксперименты, подтверждающие устойчивость модели к шуму в экспериментальных данных, обусловленному возможными ошибками в процессе их получения, а также к ограничениям выборки. Кроме того, модель продемонстрировала высокую точность и скорость вычислений по сравнению с существующими математическими моделями.

Выводы: Наш результат демонстрирует возможности, которые открываются для исследования микроэволюционных процессов и их влияния на разнообразие организмов в результате объединения компьютерных моделей популяционных процессов и модели нейронной сети. Этот подход будет полезным инструментом для дальнейших исследований в этой области.

Application of machine learning methods for analyzing genetic polymorphism in populations

Maryanovskaya T.A.[1, 2]*, Sherbakov D.Yu.[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Limnological Institut, SB RAS, Irkutsk, Russia*
* *t.maryanovskaya@g.nsu.ru*

Key words: microevolution; population genetics; deep learning

Motivation and Aim: Organisms evolve within communities of different species, and many models have been developed within the discipline of population genetics to study microevolution. The application of deep learning methods is becoming increasingly relevant in biology due to the large volume of accumulated genetic data. Microsatellite analysis poses certain challenges due to their complex structure characterized by variable lengths and repetitive nature. Studying and analyzing microsatellites require special methods and approaches for interpreting the data obtained in genetic research. The application of deep learning in population genetics opens up new possibilities for genetic data analysis and for analyzing microevolutionary processes. The objective of this work is to investigate the population dynamics in organisms with mixed reproductive strategies using deep learning methods.
Methods and Algorithms: Stochastic modeling was utilized in this study. The model specification involves a convolutional neural network with 6 internal and 2 external layers. Training of this model was carried out on data obtained from a modified Wright-Fisher model, incorporating additional asexual reproduction.
Results: This work presents a model that represents one of the first attempts to apply deep learning for tasks in population genetics. The model analyzes experimental data and predicts the ratio of sexual to asexual reproduction in organisms capable of altering their reproductive strategy, such as hydras, some crustaceans, mollusks, and others. Computational experiments confirmed the model's robustness to noise in experimental data, stemming from potential errors in their acquisition, as well as to sampling constraints. Furthermore, the model demonstrated high accuracy and computational speed compared to existing mathematical models.
Conclusion: Our results demonstrate the opportunities offered for studying microevolutionary processes and their impact on organism diversity through the integration of computer models of population processes and neural network models. This approach serves as a valuable tool for further research in this field.

doi 10.18699/bgrs2024-2.2-08

Анализ и моделирование биологических цифровых информационных систем

Ратушняк А.С.*, Проскура А.Л., Шевырина В.А.
Федеральный исследовательский центр информационных и вычислительных технологий, Новосибирск, Россия
** ratushniak.alex@gmail.com*

Ключевые слова: моделирование возникновения живых систем; энтропия; негэнтропия; молекулярные машины

Мотивация и цель: Для продвижения работ в области бионейроинформатики, для разработки методов и подходов к профилактике, диагностике и коррекции нейроиндуцированных патологий необходимо понимание базовых физических функций живых систем. Для этого необходимы представления о физической сущности и механизмах возникновения и эволюции негэнтропийных молекулярных комплексов. Работа посвящена, прежде всего, анализу существующих методов, поиску подходов к моделированию возникновения, эволюции, созданию моделей первичных протобиологических молекулярных комплексов (protobionts) [https://en.wiktionary.org/wiki/protobiont]. Анализ фундаментальных основ жизни (поведения) возможен только на основе моделирования простейшей живой системы [1, 2].
Методы и алгоритмы: Сейчас одним из ведущих постулатов считается, что основной функцией живых систем является возможность саморепликации. При этом существует понимание, что вероятность спонтанного возникновения самореплицирующейся системы беспредельно низка. А без репликации (с некоторыми ошибками в пределах порога Эйгена) эволюционный процесс, в рамках таких представлений, практически мало реален.
В рамках другого подхода к возникновению protobionts существуют представления о том, что базовой функцией живых систем является поддержание собственной устойчивости в среде после спонтанного формирования достаточно простых молекулярных конструкций. Такие конструкции, состоящие из молекул, способных выполнять логические операции, могли бы на основе информационных процессов прогнозировать состояние окружающей среды (опережающее отражение [3]) и избегать понижающих устойчивость воздействий и/или приобретать дополнительную устойчивость. Додарвиновская эволюция при этом может состоять в отборе средой наиболее устойчивых из спонтанно возникающих protobionts.
Для определения реальных путей возникновения и функционирования биосистем необходимо создание их имитационных моделей. Это облегчает анализ результатов и позволяет решать одну из важнейших задач – определение базовых функций биосистем. Однако при попытках создания моделей даже самых простейших из существующих биообъектов возникают значительные проблемы. Одной из ведущих проблем является «проклятие размерности», т. е. многокомпонентность модели. Не менее значима оценка параметров и интеграция данных. Особенно при попытках создания моделей внутриклеточных информационных сетей. Отсутствие исчерпывающих и количественных данных и потребность в вычислитель-

ных ресурсах еще больше усложняют процесс моделирования. При существующем подходе не удается создать достаточно полную модель даже одной клетки, состоящей из примерно 10^{15} атомов, при этом реакции, к примеру, сворачивания белка и ферментативные реакции происходят за пико- и фемтосекунды. Самые крупные из существующих модельных систем, доступных для используемых методов, состоят из 10^{10} атомов [2, 4], при этом временные диапазоны, в которых возможно изучение таких систем с помощью суперкомпьютеров, ограничены наносекундами. Чтобы смоделировать клетку целиком, нужна модель в 10^5 раз больше, которая, соответственно, будет работать в 10^6 раз дольше. Нужная компьютерная модель должна быть как минимум в 10^{11} раз масштабнее, чем те, что у нас есть на сегодняшний день. При существующем подходе понятно, что набор свойств объекта моделирования на порядки больше, чем свойств модели. В обозримом будущем радикального развития вычислительных мощностей не ожидается. Поэтому для решения огромного множества вопросов необходим поиск новых подходов и концепций. Наше исследование сосредоточено на решении этих проблем – выборе методов и алгоритмов моделирования биологических цифровых информационных систем.

Результаты: В данной работе мы попытались создать модель наипростейших молекулярных конструкций (агентов – негэнтропийных молекулярных комплексов), отвечающих представлениям о живых системах как об устройствах, обладающих активной устойчивостью в среде. При этом задавались параметры среды и агента. «Правила» для внешней среды: Уровень «энтропии». Наличие достаточного количества и разнообразия молекулярных компонентов для спонтанного формирования агентов, с определенной вероятностью. Наличие значимых и ассоциированных (или причинно-следственно), связанных со значимыми, подпороговых информационных сигналов (разной модальности, например, свет, звук и др.). При этом задается соотношение количества значимых и подпороговых сигналов. В среде (исходно) количество пар значимых и подпороговых сигналов может быть равно или больше, чем рецепторов у агента.

«Правила» для агентов: Агенты возникают в среде спонтанно. Задаются следующие параметры: nFr – рецепторы значимых сигналов (n – к-во); nfr – рецепторы ассоциированных сигналов (n – к-во); nE – эффекторы (n – к-во); M – объем ассоциативной памяти совпадения значимых и подпороговых сигналов; nL – запас энергии (к-во жизней) (n>1); x – коэффициент вычитания из L при срабатывании fr; y – коэффициент количества повторений f, не сочетанного с F, до удаления из памяти F+ f=E; z – максимум nL; v – коэффициент увеличения к-ва F и f при достижении L значения z. При каждом возникновении агентов значение коэффициентов обновляется (или сохраняется – наследуется).

Предельно простые супрамолекулярные конструкции флуктуационно формируются в природе, вероятно, с некоторым «запасом», избытком упорядоченности – негэнтропии N. Взаимодействие с процессами, происходящими в окружающей энтропийной среде, приводило к уменьшению этой упорядоченности, суммарной негэнтропии в системе. Приобретенная при этом информация в дальнейшем «используется» молекулярной системой для прогнозирования подобных событий на время Δt. Благодаря этому происходит уменьшение энтропии с некоторым «информационным коэффициентом» Ki, зависящим от объема полученной информа-

ции и определяющим глубину возможного прогнозирования. Разработаны алгоритмы функционирования и программные коды моделей негэнтропийных биоагентов.
Выводы: Опираясь на модель биоагента (protobionts), основанную на таких физических представлениях, и учитывая преемственность, эмерджéнтность организации биологических систем, в дальнейшем представляется возможным попытаться моделировать их эволюцию наращивать сложность до пределов аппаратных ограничений.
Финансирование: Работа выполнена в рамках государственного задания Минобрнауки России для Федерального исследовательского центра информационных и вычислительных технологий.

Analysis and modeling of biological digital information systems

Ratushnyak A.S.*, Proskura A.L., Shevyrina V.A.
Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia
** ratushniak.alex@gmail.com*

Ключевые слова: modelling emergence of living systems; entropy; negentropy; molecular machines

Motivation and Aim: To advance work in the field of bioneuroinformatics, to develop methods and approaches to the prevention, diagnosis and correction of neuro-induced pathologies, it is necessary to understand the basic physical functions of living systems. This requires ideas about the physical essence and mechanisms of the emergence and evolution of negentropic molecular complexes. The work is devoted, first of all, to the analysis of existing methods, the search for approaches to modeling the emergence, evolution, and creation of models of primary protobiological molecular complexes (protobionts) [https://en.wiktionary.org/wiki/protobiont]. Analysis of the fundamental principles of life (behavior) is possible only on the basis of modeling the simplest living system [1, 2].
Methods and Algorithms: Now one of the leading postulates is that the main function of living systems is the ability to self-replicate. At the same time, there is an understanding that the probability of spontaneous emergence of a self-replicating system is infinitely low. And without replication (with some errors within the Eigen threshold), the evolutionary process, within the framework of such ideas, is practically unrealistic.
Another approach to the emergence of protobionts is the idea that the basic function of living systems is to maintain their own stability in the environment after the spontaneous formation of fairly simple molecular structures. Such structures, consisting of molecules capable of performing logical operations, could be able, based on information processes, to predict the state of the environment (anticipatory reflection [3]) and avoid impacts that reduce stability and/or acquire additional stability. Before Darwinian evolution, it may consist in the selection by the environment of the most stable of spontaneously arising protobionts. To determine the real ways in which biosystems arise and function, it is necessary to create their simulation models. This facilitates the analysis of the results and allows us to solve one of the most important tasks – determining the basic functions of biosystems. However, when trying to create models of even the simplest existing biological objects, significant problems arise. One of the leading problems is the “curse of dimensionality”, i. e. multicomponent model. Equally important is parameter estimation

and data integration. Especially when trying to create models of intracellular information networks. The lack of comprehensive and quantitative data and, at the same time, the need for computing resources further complicate the modeling process. With the existing approach, it is not possible to create a sufficiently complete model of even one cell consisting of approximately 10^{15} atoms, while reactions, for example, protein folding and enzymatic reactions, occur in pico and femto seconds. The largest model systems existing today, available for calculation methods, consist of 10^{10} atoms [2, 4], while the time ranges in which it is possible to study such systems using supercomputers are limited to nanoseconds. To simulate an entire cell, you need a model 10^5 times larger, and which, accordingly, will work 10^6 times longer. The required computer model would have to be at least 10^{11} times larger than what we have today. With the existing approach, it is clear that the set of properties of the modeling object is orders of magnitude larger than the properties of the model. Radical developments in computing power are not expected in the near future. Therefore, to solve a huge variety of issues, it is necessary to search for new approaches and concepts. Our research focuses on addressing these issues. Selection of methods and algorithms for modeling biological digital information systems.

Results: In this work, we tried to create a model of the simplest molecular structures (agents – negentropic molecular complexes), corresponding to the concept of living systems as devices with active stability in the environment. At the same time, the parameters of the environment and the agent were set.

"Rules" for the external environment: Level of "entropy". The presence of a sufficient number and variety of molecular components for the spontaneous formation of agents, with a certain probability. The presence of significant and associated (or causally) related to significant, subthreshold information signals (of different modalities, for example light, sound, etc.). In this case, the ratio of the number of significant and subthreshold signals is set. In the environment (initially), the number of pairs of significant and subthreshold signals may be equal to or greater than the number of receptors on the agent.

"Rules" for agents: Agents arise spontaneously in the environment. The following parameters are set: nFr – receptors of significant signals (n – number); nfr – receptors of associated signals (n – number); nE – effectors (n – number); M is the volume of associative memory for the coincidence of significant and subthreshold signals; nL – energy reserve (number of lives) (n>1); x – subtraction coefficient from L when fr is triggered; y – coefficient of the number of repetitions of f not combined with F before deleting F+f=E from memory; z – maximum nL; v is the coefficient of increase in the quantity F and f when L reaches the value z. With each occurrence of agents, the value of the coefficients is updated (or saved – inherited)

Extremely simple supramolecular structures are formed fluctuation ally in nature, probably with some "reserve", an excess of order – negentropy N. Interaction with processes occurring in the surrounding entropic environment led to a decrease in this order, the total negentropy in the system. The information acquired in this case is subsequently "used" by the molecular system to predict similar events at time Δt. Due to this, entropy decreases with a certain "information coefficient" Ki, which depends on the amount of information received and determines the depth of possible prediction. Functioning algorithms and program codes for models of negentropy bioagents have been developed.

Conclusion: Based on the model of bioagents (protobionts) based on such physical concepts and taking into account the continuity and emergence of the organization of biological systems, in the future it seems possible to try to simulate their evolution and increase complexity to the limits of hardware limitations.

Funding: The research was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation for Federal Research Center for Information and Computational Technologies.

Список литературы/References

1. Thornburg Z.R., Bianchi D.M., Brier T.A. et al. Fundamental behaviors emerge from simulations of a living minimal cell. *Cell*. 2022;185(2):345-360.e28. doi 10.1016/j.cell.2021.12.025
2. Singharoy A., Maffeo C., Delgado-Magnero K.H. et al. Atoms to Phenotypes: Molecular Design Principles of Cellular Energy Metabolism. *Cell*. 2019;179(5):1098-1111.e23. doi 10.1016/j.cell.2019.10.021
3. Anokhin P.K. Anticipatory reflection of reality. *Problems of Philosophy*. 1962;7:97-106 (in Russian)
4. Karr J.R., Sanghvi J.C., Macklin D.N. et al. A whole-cell computational model predicts phenotype from genotype. *Cell*. 2012;150(2):389-401. doi 10.1016/j.cell.2012.05.044

doi 10.18699/bgrs2024-2.2-09

Об уточнении понятия адаптивности

Суслов В.В.

Институт цитологии и генетики СО РАН, Отдел системной биологии, Новосибирск, Россия
valya@bionet.nsc.ru

Key words: информация; эволюция; адаптация; отбор; сложность

Использование свойств информации в теории дарвиновой эволюции – это логически следующий шаг после формализации организмов, как систем управления (СУ) А.А. Ляпунова, выполненной В.А. Ратнером на базе расшифровки механизмов матричного воспроизводства генетических макромолекул, обеспечивающих функционирование молекулярно-генетических систем управления (МГСУ). Все теории эволюции базируются на одном, эмпирически найденном еще Ч. Дарвином и А.Р. Уоллесом свойстве, – полезности информации. В настоящей работе, используя формализацию В.И.Корогодина, к полезности добавлена нужность информации.

Не касаясь дискуссий о том, что такое информация (*I*), за точку отсчета примем то, что не оспаривается: *I* – это инструкция, записанная на носителе. В живых системах носитель – макромолекулы с матричным воспроизведением: ДНК, РНК. Такие молекулы способны к записи и передаче информации через процессы репликации, транскрипции (и обратной транскрипции) и реализации генетической информации через матричный процесс трансляции, чему в теории информации соответствует построение признака Q по информации I [1, 2]. Рождение новой информации I, зафиксированной на носителе (геноме) возникает в результате мутации. Отбор информацию не создает, а лишь оценивает [1]. Реализация информации, позволяющая оценить ее полезность, осуществляется при формировании признака Q.

Фактически, Q определенным конечным множеством способов (путей, процессов) направляет ресурсы R среды s на выживание носителя информации I (генома). Это отражено в репродуктивном вкладе P. Фишера ω [3] как оценка адаптивности Q и определяющей его информации *I* [1,2].

Согласно В.И. Корогодину [2], P – это вероятность использования ресурса R для достижения цели Z (*R→Z)* с помощью признака Q(I). Если вероятность P больше некоторой пороговой величины (P>p), то информация I полезна для достижения цели Z. В.И. Корогодин [2] формализовал полезность как $[R, s] \frac{Q(I)}{p,P} \rightarrow [Z, w]$, где w – побочный продукт достижения цели Z с использованием ресурса R. С этой точки зрения приспособленность Фишера ω – не единственная мера адаптивности, а лишь характеристика, задающая граничное условие адаптивности ω>0.

В.И. Корогодин также ввел для информации I понятие нужности, которую обозначим как U. Нужность определяется как количество (или вероятность) обращений к данной информации I [2] при условии, что имеется множество путей достижения одной и той же цели Z. Такие пути возникают в результате мутаций, изменяющих нуклеотидные последовательности. Так, прокариотический белок может быть синтезирован с гена, являющегося нормой G_0 в популяции, или с минорного варианта этого гена G* в популяции. При равной полезности генетических инструкций

G_0 и G*, нужность U максимальна для нормы G_0 (так как инструкция G_0 используется существенно чаще), а нужность U минорного варианта ниже. Учет нужности в адаптивной эволюции может привести к неожиданному результату – фиксации исходно низкополезного гена в популяции. В стационарных условиях среды высокий уровень полезности коррелирует с высокой величиной нужности. Чего нельзя сказать об изменяющейся среде.

Настоящая работа посвящена уточнению ключевого для дарвинизма понятия адаптивности путем изучения взаимосвязи между такими характеристиками генетических инструкций, как полезность и нужность.

Ранее было проведено компьютерное моделирование эволюционной динамики трофических сетей, формируемых консорциумом популяций $\{P_n\}$ прокариот [4]. Прокариоты находятся в среде, через которую идет стационарный проток неспецифического ресурса R из внешнего источника. Неспецифический ресурс равно нужен для выживания любой популяции консорциума. Кроме того, каждая популяция потребляет специфический трофический ресурс R_n. Каждый такой специфический ресурс R_n представляет собой побочные метаболические продукты w_m жизнедеятельности других популяций консорциума $\{P_n\}$. При этом консорциум характеризуется наличием трофических сетей. Каждая трофическая сеть представляет собой несколько популяций, связанных посредством множества метаболических продуктов $\{w_m\}$: одни популяции нарабатывают и выделяют в среду определенные метаболиты, а другие их потребляют (и наоборот). В работе [4] с использованием программного продукта «Эволюционный конструктор», предназначенного для моделирования эволюции популяций гаплоидных организмов, было показано следующее. (1) При варьировании поступления в консорциум неспецифического ресурса R происходил распад ряда прежних трофических сетей и образование новых. (2) Критически важным для выживания конкретной популяции $\{P_n\}$, характеризующейся определенным генотипом, являлось количество трофических подсетей, в которые включалась данная популяция посредством потребления различных метаболитов w_n. Согласно результатам работы [4], параметр ω (репродуктивный вклад в следующее поколение популяции, определяемый по Р.Фишеру [3]), оказался не оценкой адаптивности (т.е. способности популяции выживать, как принято в синтетической и эпигенетической теориях эволюции (СТЭ и ЭТЭ) [3]), а лишь ее граничным условием (ω>0 для невырождающихся популяций).

Параметр ω оказался граничным условием адаптивности и в долговременном (более 20 лет) эволюционном эксперименте с популяциями *E. coli* [5]. Выявление мутаций в генотипах и оценка ω для таких генотипов осуществлялись по изменению скорости роста популяций *E. coli*. Мутантные штаммы в этом эксперименте были основателями новых популяций. Для них в каждом поколении оценивалась адаптивность по Р. Фишеру (ω), характеризующая их репродуктивный вклад в следующее поколение. Оказалось, что мутации, сами по себе ведущие к падению ω, могли открывать новые возможности для эволюции в компаунде с другими мутациями, что и реализовывалось в следующих поколениях [5].

Описанный долговременный эксперимент [5], как показал проведенный анализ литературы, подтверждает полученные более 30 лет назад результаты компьютерного анализа, проведенного Ю.В. Чайковским [6], который изучал эволюционную динамику модельных клонов, чья адаптивность ω оценивалась по Фишеру.

В недавно выполненном цикле работ [7, 8] исследована эволюция активности стероидных рецепторов в той области их белковой молекулы, которая отвечает за

связывание стероидного гормона. Изучалась эволюция этих рецепторов в процессе филогенеза позвоночных [7], включая: (а) построение филогенетических деревьев и реконструкцию предковых молекул и (б) синтез генов, а также наработка с них соответствующих белков. Замечательный результат заключался в том, что были выявлены мутации, снижающие активность (или стабильность) белков-стероидных рецепторов, но в компаунде с другими мутациями, резко повышающих эти характеристики.

В работе [8] этот же результат был получен на другой основе. Описанные выше синтезированные гены стероидных рецепторов были встроены в геномы дрожжей, для которых был проведен отбор на усиление активности стероидных рецепторов по связыванию стероидного гормона.

Характерная особенность нового режима молекулярной эволюции, названного авторами работ [7, 8] «конформационным эпистазом» заключался в следующем. В процесс эволюции могут возникать и фиксироваться мутации, которые: (а) снижают специфическую активность белков, величина которой ранее являлась адаптивным признаком организма, но (б) дают возможность фиксации других замен уже как адаптивных.

Приведенные результаты свидетельствуют о необходимости дальнейшего развития теории адаптивной эволюции, учитывающей не только репродуктивный вклад ω в следующее поколение, но также другие характеристики генетической информации.

Финансирование: FWNR-2022-0020.

References

1. Чернавский Д.С. Синэргетика и информация. LENAND, URSS. 2017. 304 с.
2. Корогодин В.И., Корогодина В.Л. Информация как феномен жизни. В: Корогодин В.И. Феномен жизни. Т. 2. М.: Наука, 2012:173-338
3. Herron J., Freeman S. Evolutionary Analysis. Pearson. 2014. 864 p.
4. Lashin S.A. et al. Simulation of coevolution in community by using the "Evolutionary Constructor" program. *In Silico Biol.* 2007;7(3):261-275
5. Woods R.J. et al. Second-order selection for evolvability in a large *Escherichia coli* population. *Science*. 2011;331(6023):1433-1436
6. Чайковский Ю.В. Выживание мутантного клона. Сообщ. 1. Сообщ. 2. *Генетика*. 1977; 13(8):1467-1477, 1478-1488
7. Ortlund E.A. et al. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science*. 2007;317(5844):1544-1548
8. Starr T.N. et al. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature*. 2017;549(7672):409-413

doi 10.18699/bgrs2024-2.2-10

Моделирование фенологического отклика травянистых растений на изменение климата в Западной Сибири

Фомин Э.[1]*, Фомина Т.[2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Центральный сибирский ботанический сад СО РАН, Новосибирск, Россия*

* *fomin@bionet.nsc.ru*

Ключевые слова: модели фенологического развития; метеорологические и фенологические тренды; травянистые однолетники; Западная Сибирь

Мотивация и цель: Фенология растений является одним из надежных биоиндикаторов глобального изменения климата [1]. Многочисленные исследования в умеренных экосистемах Северного полушария согласуются с ожидаемыми смещениями фаз развития в соответствии с трендами изменения температуры [2]. При этом отмечается не только высокая вариабельность смещений фенологии в зависимости от региона, видового состава и прочих факторов [3], но и различная чувствительность видов и высокая контрастность реакций даже в пределах одного местообитания [4, 5].

Доступные в Интернете базы данных фенологических событий из-за многообразия сезонных проявлений развития и зачастую слабой их выраженности не имеют достаточной полноты и охватывают преимущественно фрагментарные наблюдения за отдельные годы, интересные конкретным исследователям и коллективам. Такие базы данных подвержены человеческому фактору, имеют пропуски, случайные и систематические ошибки регистрации дат фенособытий, преимущественно в сторону запаздывания. Необходимые для анализа метеорологические данные в основном собираются приборами и существенно меньше подвержены человеческому фактору, хотя пространственная разделенность мест регистрации метео- и фенонаблюдений также вносит множество случайных ошибок.

Недостаточность фенологических данных не позволяет обучать нейронные сети общего назначения в силу переизбытка внутренних параметров относительно числа входных данных, и необходимо использовать более простые модели. Целью нашей работы является обработка накопленных за более чем 20-летний период данных наблюдений за фенологией травянистых многолетников и разработка малопараметрических моделей для прогноза фенологического развития растений.

Методы и алгоритмы: Фенологические наблюдения проводились в коллекции декоративных растений природной флоры Центрального сибирского ботанического сада (ЦСБС СО РАН) на 78 видах многолетников (55°2′29.4″ с.ш., 82°56′4.6″ в.д.). Массовый сбор данных осуществлялся в период 1996–2015 гг. Фенонаблюдения проводили регулярно, 2–3 раза в неделю, в течение всего сезона вегетации. Число особей разных видов различалось по годам, но на момент наблюдения было не менее 4–6 шт. особей зрелого генеративного состояния каждого вида.

Данные метеорологических (метеостанция Огурцово № 29638) и фенологических наблюдений обрабатывались стандартными методами с помощью пакетов PAST, MSExcel, R и методом бутстрэппинга [6], реализованного собственным кодом на C++. Для разработки прогностических моделей развития используется Python. В основе таких моделей лежит короткая цепочка итерационных персептронов, обучаемых на фенологических базах данных небольшого объема (рис. 1).

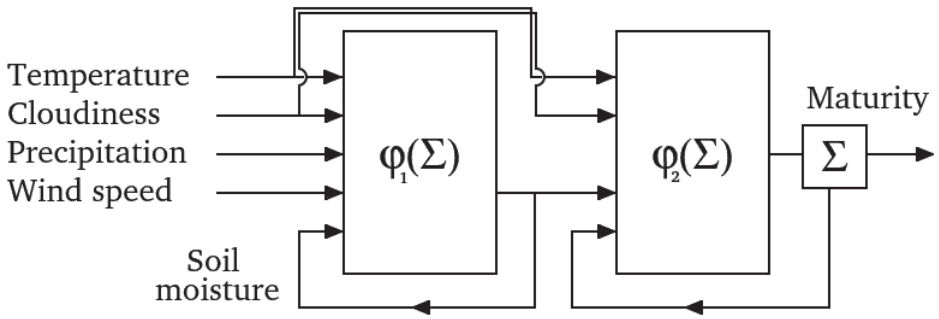


Рис. 1. Пример простейшей модели для прогноза одиночной фенофазы развития (φ_2). Модель дополнена персептроном (φ_1) для подмены параметра «осадки» на параметр «влажность почвы»

Результаты: Обработаны данные фенологических наблюдений 78 видов многолетников коллекции декоративных растений природной флоры ЦСБС СО РАН [7]. Получены результаты, указывающие направление изменений фенособытий для разных экологических групп [8] и разных жизненных форм [9] в условиях изменения климата для Западной Сибири. Для более точного и видового прогноза фенособытий «в режиме реального времени» предложена модель, основанная на цепи связанных итерационных персептронов. Модель позволяет подменять типы сумматоров (Σ) и функции активации (φ)? компенсируя негибкость малопараметрических подходов. Один из возможных вариантов модели, построенный с использованием обобщенной линейной функции (φ) в условиях экспоненциального и линейного роста (Σ)? опробован на данных статьи [10]. Получено, что модель качественно правильно прогнозирует отсутствие влияния осадков и влажности воздуха на урожайность риса («обнуление» связанных с данными параметрами входных данных) и рост/снижение урожайности с изменением продолжительности сезона.

Выводы: Разработаны программы для получения метеорологических и фенологических трендов и оценки их достоверности на основе метода бутстрэппинга. Предложены прогностические модели, способные обучаться на основе фенологических баз данных небольшого объема.

Финансирование: Исследование поддержано бюджетными проектами № FWNR-2022-0020 и АААА-А21-121011290025-2. При подготовке статьи использовались материалы Биоресурсной научной коллекции ЦСБС СО РАН, УНУ «Коллекции живых растений в открытом и закрытом грунте», USU 44053.

Modeling the phenological response of herbaceous plants to climate change in Western Siberia

Fomin E.[1]*, Fomina T.[2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Central Siberian Botanical Garden, SB RAS, Novosibirsk, Russia*

* *fomin@bionet.nsc.ru*

Key words: models of phenological development; meteorological and phenological trends; herbaceous annuals; Western Siberia

Motivation and Aim: Plant phenology is one of the reliable bioindicators of global climate change [1]. Numerous studies in temperate ecosystems of the Northern Hemisphere are consistent with expected shifts in developmental phases in accordance with temperature trends [2]. At the same time, there is not only high variability in phenology shifts depending on the region, species composition and other factors [3], but also different sensitivity of species and high contrast of reactions even within the same habitat [4, 5]. Databases of phenological events available on the Internet, due to the diversity of seasonal manifestations of development and often their weak expression, are not sufficiently complete and cover mainly fragmentary observations for individual years that are interesting to specific researchers and teams. Such databases are subject to the human factor, have omissions, random and systematic errors in recording the dates of phenological events, mainly in the direction of lag. The meteorological data required for analysis is mainly collected by instruments and is significantly less susceptible to human error, although the spatial separation of places where meteorological and phenological observations are recorded also introduces many random errors.

The insufficiency of phenological data does not allow training general-purpose neural networks due to an overabundance of internal parameters relative to the number of input data, and it is necessary to use simpler models. The goal of our work is to process observational data accumulated over more than 20 years on the phenology of herbaceous perennials, and to develop low-parameter models for predicting the phenological development of plants.

Methods and Algorithms: Phenological observations were carried out in the collection of ornamental plants of the natural flora of the Central Siberian Botanical Garden (CSBS SB RAS) on 78 species of perennials (55°2′29.4″ N, 82°56′4.6″ E). Bulk data collection was carried out during the period 1996–2015. Phenomena observations were carried out regularly, 2–3 times a week, throughout the growing season. The number of individuals of different species varied by year, but at the time of observation there were at least 4–6 individuals in a mature generative state of each species. Data from meteorological (Ogurtsovo weather station No. 29638) and phenological observations were processed using standard methods using the PAST, MSExcel, R packages and the bootstrapping method [6], implemented with our own code in C++. Python was also used to develop development models. The predictive models currently being developed are based on a short chain of iterative perceptrons trained on small phenological databases (Fig. 1).

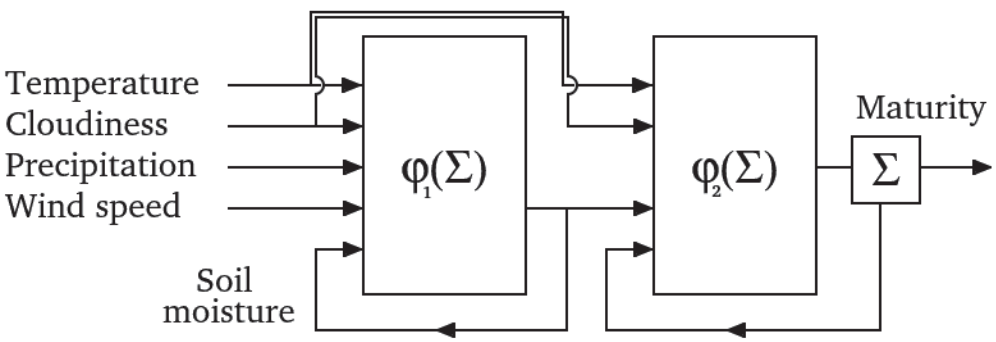


Fig. 1. An example of the simplest model for predicting a single phenophase of development (φ_2). The model is supplemented with a perceptron (φ_1) to replace the "precipitation" parameter with the "soil moisture" parameter

Results: Text Data from phenological observations of 78 species of perennials from the collection of ornamental plants of the natural flora of the Central Botanical Garden of the Siberian Branch of the Russian Academy of Sciences were processed [7]. Results were obtained indicating the direction of changes in pheno-events for different ecological groups [8] and different life forms [9] under climate change conditions for Western Siberia.For a more accurate and specific prediction of pheno-events "in real time", a model based on a chain of connected iterative perceptrons is proposed. The model allows you to change the types of adders (Σ) and activation functions (φ), compensating for the inflexibility of low-parameter approaches. One of the possible variants of the model, constructed using a generalized linear function (φ) under conditions of exponential and linear growth (Σ), was tested on the data of the article [10]. It was found that the model qualitatively correctly predicts the absence of the influence of precipitation and air humidity on rice yield ("zeroing" the input data associated with these parameters) and the increase/decrease in yield with changes in the length of the season.
Conclusion: Programs have been developed to obtain meteorological and phenological trends and assess their reliability based on the bootstrapping method. Predictive models are proposed that can be trained on the basis of small phenological databases.
Funding: The study was supported by budget projects No. FWNR-2022-0020 and AAAA-A21-121011290025-2. In preparing the article, materials from the Bioresource Scientific Collection of the Central Botanical Garden of the Siberian Branch of the Russian Academy of Sciences, UNU "Collections of living plants in open and closed ground", USU 44053 were used.

Список литературы/References

1. Lippmann R., Babben S., Menger A., Delker C., Quint M. Development of wild and cultivated plants under global warming conditions. *Curr Biol.* 2019;29:1326-1338. doi 10.1016/j.cub.2019.10.016
2. Cleland E.E., Chuine I., Menzel A., Mooney H.A., Schwartz M.D. Shifting plant phenology in response to global change. *Trends Ecol Evol.* 2007;22(7):357-365. doi 10.1016/j.tree.2007.04.003
3. Piao S., Liu Q., Chen A., Janssens I.A., Fu Y., Dai J., Liu L., Lian X., Shen M., Zhu X. Plant phenology and global climate change: Current progresses and challenges. *Glob Chang Biol.* 2019;25:1922-1940. doi 10.1111/gcb.14619
4. Lehoczky A., Szabó B., Pongrácz R., Szentkirályi F. Testing plant phenophase as proxy: Sensitivity analysis of first flowering data from the 19th century. *Appl Ecol Environ Res.* 2016;14(2):213-233. doi 10.15666/aeer/1402_213233
5. Chmura H.E., Kharouba H.M., Ashander J., Ehlman S.M., Rivest E.B., Yang L.H. The mechanisms of phenology: the patterns and processes of phenological shifts. *Ecol Monogr.* 2019;89(1):e01337. doi 10.1002/ecm.1337
6. Efron B. Bootstrap methods: Another look at the jack-knife. *Ann Stat.* 1979;7(1):1-26. doi 10.1214/aos/1176344552
7. Fomin E.S., Fomina T.I. Changes in the Phenology of Perennial Plants in Western Siberia against the Background of Global Warming. *Contemp Probl Ecol.* 2021;14(5):434-445. doi 10.1134/S199542552105005X
8. Fomin E.S., Fomina T.I. Phenological Reactions of Perennial Plants to Climate Change in Western Siberia. *Contemp Probl Ecol.* 2023;16(6):698-707. doi 10.1134/S1995425523060070
9. Fomin E.S., Fomina T.I. Phenological Response of Plants of Different Biomorphs to Climate Change in Western Siberia. *Biol Bull.* 2024;51(3):565-575. doi 10.1134/S1062359023604445
10. Rebollo I., Scheffel S., Blanco P., Molina F., Martínez S., Carracelas G., Aguilar I., de Vida F.P., Rosas J.E. Consolidating 23 years of historical data from an irrigated subtropical rice breeding program in Uruguay. *Crop Sci.* 2023;63:1300-1315. doi 10.1002/csc2.20955

doi 10.18699/bgrs2024-2.2-11

Modeling bistable dynamics of bacterial restriction-modification systems to understand bacterial defense systems

Djordjevic M.[1], Zivkovic L.[1], Ou H.-Y.[2], Djordjevic M.[3]*

[1] *Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia*

[2] *Molecular Microbiology Laboratory, Shanghai Jiaotong University, Shanghai, China*

[3] *Quantitative Biology Group, University of Belgrade – Faculty of Biology, Belgrade, Serbia*

* *dmarko@bio.bg.ac.rs*

Key words: gene expression regulation; bistability; biophysical models; quantitative biology; pathogenic gene dissemination; bacteriophages

Motivation and Aim: Type II restriction-modification (R-M) systems, crucial bacterial defenses against invasive genetic elements, are the focus of this study [1]. We introduce a novel mathematical model that delves into the dynamic interplay between restriction enzymes (R) and methyltransferase (M) within these systems governed by a regulatory protein C [2]. This model is a significant step towards unraveling the complexities of R-M systems and predicting their behavior under varying biological conditions.

Methods and Algorithms: We constructed a detailed mathematical model incorporating biophysical evidence to minimize reliance on arbitrary parameters [3]. This model includes the transcription dynamics of the control protein C and its feedback mechanisms, which significantly affect R and M protein levels. Stability analysis and bifurcation diagrams were used to explore the conditions leading to monostability and bistability, which are crucial for understanding how these systems respond to environmental changes. Numerical simulations complemented analytical derivations to validate the model against experimental data.

Results: Our model accurately predicted the observed variations in the M-to-R ratios across different R-M systems, such as Esp1396I, AhdI, and EcoRV, which correspond to distinct regulatory mechanisms and parameters [3]. The results demonstrated that changes in system dynamics, including shifts from monostability to bistability, significantly affect the susceptibility of bacteria to phage infections and the effectiveness of the bacterial defense strategy. The model also successfully captured the impact of external factors like plasmid copy number and growth rates on the M-to-R ratio.

Conclusion: This study presents a comprehensive theoretical framework that deepens our understanding of R-M system dynamics, with direct implications for bacterial adaptability and evolution [4]. It underscores the critical role of the regulatory protein C in modulating the balance between restriction and modification enzymes, thereby shaping the bacterial defense mechanisms against HGT. This framework not only provides insights into the biophysical underpinnings of these systems but also paves the way for future experimental designs and therapeutic strategies to combat the spread of antibiotic resistance.

Funding: This work is supported by The Science Fund of the Republic of Serbia (Grant no. 7750294, q-bioBDS). The study is also supported by the National Natural Science Foundation of China, Grant no. 32370186.

References

1. Mruk I., Kobayashi I. To be or not to be: regulation of restriction–modification systems and other toxin–antitoxin systems. *Nucleic Acids Res*. 2014;42(1):70-86
2. Kirillov A., Morozova N., Kozlova S., Polinovskaya V., Smirnov S., Khodorkovskii M., Severinov K. Cells with stochastically increased methyltransferase to restriction endonuclease ratio provide an entry for bacteriophage into protected cell population. *Nucleic Acids Res*. 2022;50(21):12355-12368
3. Djordjevic M., Zivkovic L., Ou H.-Y., Djordjevic M. Nonlinear Regulatory Dynamics of Bacterial Restriction-Modification Systems Modulates Horizontal Gene Transfer Susceptibility. Under revision. 2024
4. Kooni E.V., Makarov K.S., Wolf Y.I. Evolutionary genomics of defense systems in archaea and bacteria. *Annual Rev Microbiol*. 2017;71:233-261

doi 10.18699/bgrs2024-2.2-12

Demographics of genetic admixture and expansion

Ediev D.M.[1, 2, 3]*

[1] *Mathematics Chair, North-Caucasian State Academy, Cherkessk, Russia*

[2] *Demography Chair (HSMSS), Lomonosov Moscow State University, Moscow, Russia*

[3] *International Institute for Applied Systems Analysis, Laxenburg, Austria*

* *ediev@ncsa.ru*

Key words: genetic admixture; genetic expansion; ethnogenesis; demography

Motivation and Aim: We study the dynamics of genetic admixture and expansion, as well as language assimilation, through mathematical-demographic modelling. Our primary aim is to address the 'paradox' in population genetics, where autosomes and allosomes can provide markedly different, if not contradictory, pictures of past migrations. More broadly, we aim to develop a model setup that would allow the reconstruction of past demographic parameters that might have produced a given population genetic outcome.
Methods and Algorithms: We develop population reproduction models for each of the types of genetic markers studied: a single-sex heterogeneous population model for uniparental markers (allosomes) and two-sex heterogeneous population models for autosomes and for the linguistic assimilation model. To facilitate modeling the long-term population change without unnecessary complications due to age structural change, we use the reproductive value [1, 2] as our main indicator of contribution to the population genome. Our input variables are the initial population sizes, reproductive advantage of the initial elite and marriage selectivity parameters.
After developing the formalism for the proposed demographic models, we perform numerical simulations aimed at emulating two scenarios stylised after two empirical populations. The first empirical population is that of the Beaker Complex people of Britain [3], whose migration to Britain led, according to palaeogenetic data, to a rapid and almost complete turnover of the population genome. The second empirical population is that of the Karachay-Balkar people, two closely related Turkic-speaking North Caucasian autochthons who show a complex population genetic pattern: while their Y-DNA shows a contribution of about one third of steppe origin, their autosomal DNA is no different from that of their Caucasian neighbours and shows only a minor contribution attributable to East Turkic migrations [4–6].
Results: We develop formal relations for the processes of expansion of the allosomes and of admixture of autosomal DNA using, in the latter case, both selective and non-selective marriages scenarios. To examine consequences of social adaptation, we also develop models of linguistic assimilation similar to the models of genetic admixture. We demonstrate that the aforementioned paradox may find a purely demographic explanation, as single-sex and two-sex reproduction models exhibit markedly distinct dynamics. We illustrate that the three processes – allosomal expansion, autosomal admixture, and language assimilation – occur at significantly different modes, potentially explaining the varied outcomes of these processes upon the completion of ethnogenetic transitions.
The genetic expansion process, which describes the evolution of non-recombinant allosomes, proceeds rapidly by historical standards. Demographically, this expansion process can be effectively modeled as a single-sex heterogeneous population.

Even more rapid is the process of autosomal admixture, when intermarriages, even if limited, are permitted. This admixture, essentially modeled as a two-sex reproduction process, leads to the swift convergence of both "invaders" and "autochthones" in terms of autosomes. Consequently, the autosomal genetic makeup stabilizes quickly and becomes independent of the allosome expansion. This decoupling explains why population genetic makeups often exhibit significant divergence between allosomes and autosomes. The rapidity of genetic admixture suggests that the initial reproductive advantage (or disadvantage) of immigrants may rapidly diminish through social assimilation processes, which we modeled using linguistic assimilation formalism. By designating one language as "dominant" (either superstrate or substrate) and assuming that every child with a parent fluent in that language adopts it, our model captures the essence of social assimilation. While real-life assimilation may be slowed by cultural norms or accelerated by schooling and social interaction, our findings demonstrate that the assimilation process can occur swiftly, even faster than genetic admixture. This phenomenon of social assimilation explains why initial reproductive differentials can quickly dissipate, halting genetic expansion and leading to complex population genetic make-ups at the end of ethnogenesis.

Our numerical illustrations demonstrate that the formal models proposed offer valuable insights into reconstructing migrations and mechanisms of ethno-genesis in the past. The first illustration supports the hypothesis in the literature that the Beaker Complex people likely arrived in Britain during a period of sparse population. Simulations based on the genetic makeup of the Karachay-Balkars lend credibility to the scenario of elite dominance in their formation. Our results suggest a small (up to about five per cent) initial migration of powerful elites who had a strong reproductive advantage but freely intermarried with the local population.

Building upon formal results and numerical simulations, we propose several qualitative indicators of modes of social transformation in historical populations. A large disparity between the proportions of invading genome in allosomes and autosomes serves as an indicator of quick admixture, suggesting minimal obstacles to interethnic marriages. Conversely, similar proportions of invading genome in allosomes and autosomes indicate lack of mixed marriages. A limited eventual proportion of invading allosomes suggests quick social assimilation and ethno-genesis, highlighting the rapid integration of immigrant populations into the existing social fabric.

Conclusion: Our research offers valuable insights into the intricate interplay between demography, genetics, and social organization, providing implications for historical scenarios and enhancing our understanding of the long-term consequences of migration and social cohesion.

The fluidity of linguistic transitions, even in comparison to population genetic transitions, underscores the need for caution when employing linguistic arguments to study ethnicity over extended historical periods.

Another practical implication of our results is that while the genetic composition of the population after the expansion process is complete may show a substantial contribution from the newcomers, especially in the allosomes, their initial share of the genome may be much smaller. This is all the more true for the population share, as the per capita reproductive value of immigrants would typically be higher than that of the autochthonous population due to the younger age composition. This suggests that it may be difficult to trace the original migration events using direct palaeogenetic evidence, especially if the first generations of invaders buried their deceased in their former place of origin.

Our findings highlight the need for deeper, more nuanced population genetic analyses that do not focus solely on autosomal or allosomal genes. Analysing different aspects of a population's genetic make-up together will provide more robust conclusions and more meaningful insights into the population's past.

Last but not least, our models may be useful for better understanding the population genetic consequences of modern migrations in large cities or developed countries, where migration has increased to levels that affect population reproduction [7].

Our models can be extended in several ways to improve our understanding of past migrations: by modelling both Y and MtDNA simultaneously, by considering stochastic and diffusion effects, by assuming multiple or even continuous migrations.

Funding: This study did not receive external funding.

References

1. Fisher R.A. The genetical theory of natural selection. In: The genetical theory of natural selection. Clarendon Press, 1930. doi 10.5962/bhl.title.27468
2. Ediev D.M. On an extension of R.A. Fisher's result on the dynamics of the reproductive value. *Theor. Popul. Biol*. 2007;72(4):480-484. doi 10.1016/j.tpb.2007.03.001
3. Armit I., Reich D. The return of the Beaker folk? Rethinking migration and population change in British prehistory. *Antiquity*. 2021;95(384):1464-1477. doi 10.15184/AQY.2021.129
4. Yunusbayev B., Metspalu M., Metspalu E., Valeev A., Litvinov S., Valiev R., Akhmetova V., Balanovska E., Balanovsky O., Turdikulova S., Dalimova D., Nymadawa P., Bahmanimehr A., Sahakyan H., Tambets K., Fedorova S., Barashkov N., Khidiyatova I., Mihailov E., … Villems R. The Genetic Legacy of the Expansion of Turkic-Speaking Nomads across Eurasia. *PLoS Genet*. 2015;11(4):e1005068. doi 10.1371/JOURNAL.PGEN.1005068
5. Alexeev V.P. Origins of the peoples of the Caucasus. A craniological study. Moscow: Nauka Publ., 1974 (in Russian)
6. Djaubermezov M.A. Gene pool of the Balkar and Karachay populations according to a comprehensive study of mitochondrial DNA, Y-chromosome and whole genome analysis. Phd dissertation thesis, biological sciences. Ufa, 2019 (in Russian)
7. Ediev D.M., Coleman D., Scherbov S. New measures of population reproduction for an era of high migration. *Population, Space and Place*. 2013;20(7):622-645. doi 10.1002/psp.1799

doi 10.18699/bgrs2024-2.2-13

On the role of natural selection in the genetic divergence of migration-coupled populations: results of mathematical modeling and experiments

Frisman E.Ya.[1], Zhdanova O.L.[1, 2]*

[1] *Institute for Complex Analysis of Regional Problems, FEB RAS, Birobidzhan, Russia*

[2] *Institute of Automation and Control Processes, FEB RAS, Vladivostok, Russia*

* *axanka@iacp.dvo.ru*

Key words: evolution; population genetics; natural selection; genetic drift; mathematical modelling; polymorphism; *Drosophila melanogaster*

Motivation and Aim: The existence of stable differences in a uniform selection area seems somewhat paradoxical from the point of view of population genetics; however, it is precisely disruptive selection that can give such divergence [1]. Examples of population systems with stable divergence of genetic structures of subpopulations inhabiting a homogeneous area are found in nature and created experimentally [2–4]. Comparing these observations and modeling results allows us to explain the maintenance of the stability of natural genetic divergence by the interaction of disruptive selection (in the form of reduced fitness of hybrid forms) and weak migration process. The results of experiments performed by Yu.P. Altukhov and co-authors with populations of Drosophila melanogaster, in which primary divergence of the genetic structures of the subpopulation was obtained at the α-GDH locus, are in good agreement with this statement. One can assume that disruptive selection, in this case – reduced fitness of α-GDH heterozygotes, played a significant role in maintaining this divergence. An alternative hypothesis is the fixation of differences by genetic drift.

On the one hand, to identify and substantiate the possibility of the presence of this factor in the system, and on the other hand, to assess how significant a role disruptive selection plays in maintaining the stability of primary genetic divergence, in this work we analyzed mathematical models of the dynamics of allele frequencies in a large panmictic population and a system of 30 local migration-coupled populations, and also compared the model dynamics of allele frequencies under the action of selection and without it with the data obtained in the experiment.

Methods and Algorithms: In the experimental study [5, 6], Yu.P. Altukhov and colleagues studied the dynamics of allele frequencies of two autosomal diallelic loci α-glycerophosphate dehydrogenase (α-GDH: 2 alleles, chromosome II) and esterase-6 (Est-6: 2 alleles, chromosome III) in a system of 30 sequentially migration-related vial populations of Drosophila melanogaster. One hundred and fifty pairs of males and females heterozygous for both loci were placed in one of the compartments of the population vial and were allowed to reproduce freely and disperse throughout the area. After all compartments were populated and the number of flies stabilized (around 135 individuals), a panmictic population of the same size was founded, practically fully reproducing the gene pool of the evolutionary system. Both populations were kept in a thermostatically controlled carton and maintained under completely identical conditions, such as feeding, lighting, humidity, generation turnover, etc. The system evolved for about

70 generations. As a result of this experiment, the conditions that ensured a low migration rate and relatively random initial allele frequencies in most subpopulations, and local differentiation of allele frequencies at both loci appeared in the system despite completely identical conditions in the subpopulations. In addition, the displacement of one of the alleles of the α-GDH locus occurs, which was most noticeable in the large panmictic population.

To model the experimental dynamics of allele frequencies in the large panmictic population, we used the classic selection model in a population with non-overlapping generations, the coefficients of which were estimated from the experimental data by the least squares method. Furthermore, based on the fact that experimental results one can explain by at least two alternative hypotheses: the adaptive hypothesis, which assumes the presence of disruptive selection, and the fluctuation hypothesis, which assumes purely random drift differentiation. To assess the adequacy of these hypotheses, we used a mathematical model of the dynamics of gene frequencies in a chain of 30 migration-coupled ring populations and analyzed the results of its multiple implementations in a computer model [7].

Results and Conclusion: Based on the results of computer modeling, we can conclude that under disruptive selection, primary genetic divergence in a system of populations connected by stochastic migrations arises with a high probability and proves to be structurally stable. At the same time, comparing the experimental data with the modeling results shows that in the artificial population system constructed by Yu.P. Altukhov and colleagues, the disruptive selection contributed to the primary genetic divergence of populations.

Looking at the ongoing processes more broadly, we can obtain the following generalization. Since hybrid individuals created the original core of the experimental population system, we can assume that during the development of the system, two opposite processes took place in parallel. The first was the formation of gene blocks with increased fitness of heterozygous forms. The process of natural selection here maintained polymorphism of the alleles of these genes. The second process was associated with the segregation of heterozygous forms and the identification of genes characterized by increased fitness of homozygotes and disruptive dynamics of allele frequencies. Both of these processes involved alleles of genes that were adaptively neutral in this experimental system. For various reasons, the main of which is linkage, these neutral alleles turned out to be genetic markers of the corresponding adaptive processes. Thus, the esterase-6 locus turned out to be such a marker for processes aimed at maintaining allele polymorphism, and the α-GDH locus – a marker of disruptive selection occurring in the system and contributing to primary genetic divergence.

Funding: The study was carried out within the state assignment of IACP FEB RAS (Theme FWFW-2021-0004) and ICARP FEB RAS (Theme FWUG 2024-0005).

References

1. Bazykin A.D. Disadvantages of heterozygotes in a system of two adjacent populations. *Genetika = Genetics.* 1972;8(11):155-161 (in Russian)
2. Gordeeva N.V., Salmenkova E.A., Altukhov Y.P. Genetic divergence in pink salmon introduced into the European North of Russia: Microsatellite and allozyme variation analysis. *Russ J Genet.* 2006;42(3):268-278. doi 10.1134/S1022795406030069
3. Kandul N.P., Lukhtanov V.A., Pierce N.E. Karyotypic diversity and speciation in *Agrodiaetus butterflies. Evolution*. 2007;61(3):546-559. doi 10.1111/j.1558-5646.2007.00046.x
4. Eo S.H., DeWoody J.A. Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. *Proc Royal Soc B: Biol Sci.* 2010;277(1700):3587-3592. doi 10.1098/rspb.2010.0965

5. Altukhov Yu.P., Bernashevskaya A.G., Milishnikov A.N., Novikova T.A. Experimental modeling of genetic processes in the population system of Drosophila melanogaster corresponding to the circular stepping-stone model: I. Substantiation of the approach and characteristics of local differentiation of the α-glycerophosphate dehydrogenase and esterase-6 allele frequencies. *Genetika = Genetics.* 1979;15(4):646-653 (in Russian)
6. Altukhov Yu.P., Bernashevskaya A.G. Experimental modelling of genetic processes in a population system of *Drosophila melanogaster* corresponding to the circular stepping-stone model: 2. Stability of allelic composition and periodic relationship of allele frequency with distance. *Genetika = Genetics.* 1981;17(6):1052-1059 (in Russian)
7. Zhdanova O.L., Frisman E.Y. On the genetic divergence of migration-coupled populations: modern modeling based on the experimental results of Yu. P. Altukhov et al. *Russ J Genet*. 2023;59(6):614-622. doi 10.1134/S1022795423060133

doi 10.18699/bgrs2024-2.2-14

Integrated mathematical modeling, experimental and bioinformatics study of Type-II antitoxin-toxin system's response to antibiotic exposure

Ilic B.[1]*, Djordjevic M.[2], Ou H.-Y.[3]

[1] *Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia*

[2] *Faculty of Biology, University of Belgrade, Belgrade, Serbia*

[3] *Shanghai Jiao Tong University, Shanghai, China*

* *bojanab@ipb.ac.rs*

Key words: Type II toxin-antitoxin systems; antibiotic persistence; systems biology; bioinformatics; non-linear dynamics; gene expression regulation;

Motivation and Aim: Antibiotic persistence refers to a phenomenon in which a small subpopulation of bacterial cells (dormant cells) survives antibiotic treatment despite being genetically identical to the vast majority of cells that are susceptible to antibiotics. Bacterial toxin-antitoxin (TA) modules of Type-II are involved in response to various stressful conditions, such as antibiotic exposure. Elevated toxin levels may induce dormancy. Our goal is to explore how the kacAT Type-II TA system (a member of the GNAT-ribbon-helix-helix family) participates in the antibiotic persistence of *Klebsiella pneumoniae.*

Methods and Algorithms: We tackle this problem by joint theoretical and experimental study of kacAT gene expression regulation [1], as well as with a bioinformatics and statistical analysis of GNAT-RHH loci in completely sequenced genomes of *K. pneumoniae* [2]. To this end, and based on the fact that KacAT complex formation is cooperative, we developed a mechanistic model of system dynamics, which is characterized by highly non-linear feedback loops and a single stationary state (Fig. 1A).

Results: Analysis of our mathematical model [1] enables us to reproduce the following experimental observations: (i) Reduction of [KacA]:[KacT] protein ratio upon antibiotic application (antibiotic levels increase, i.e., antitoxin degradation increases going from left to right along each curve in Fig. 1B). (ii) Large kacAT transcript increase induced by antibiotics, which based on the model can be achieved under strong promoter autorepression by KacAT complex (Fig. 1C). (iii) KacAT overexpression induces tolerance to antibiotic stress, whereas kacAT deletion does not affect this tolerance (predicted large binding affinity Kb in Fig. 1C).

The finding that kacAT deletion does not lead to spontaneous persister formation is consistent with growing experimental evidence in different Type-II TA systems but opposes earlier theoretical models (see, e.g. [3]), which predicted the appearance of bistability. However, these models assumed the cooperative (joint) action of multiple TA systems located at several instances in the genome. So, we performed a bioinformatics analysis of GNAT-RHH family loci in completely sequenced genomes of *K. pneumoniae* [2]. Statistical analysis of thus predicted loci indicates that only one system in the same family as kacAT is strongly preferred (Fig. 2A–E). Additionally, we obtain statistically significant negative correlations between different clades for which experiments indicate their cross-talk so that such interactions are disfavored (Fig. 2F).

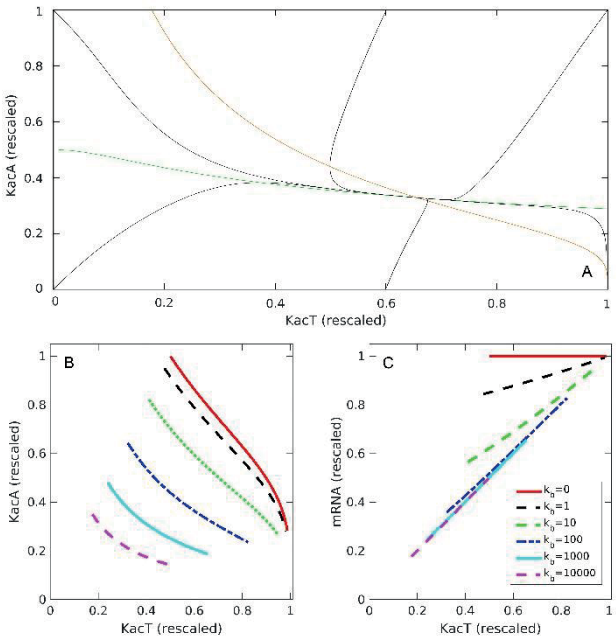


Fig. 1. Model of kacAT expression dynamics. (A) Phase plane analysis. (B) Equilibrium values of proteins KacA *vs* KacT. (C) Transcript kacAT mRNA *vs* KacT. Figure adapted from [1]

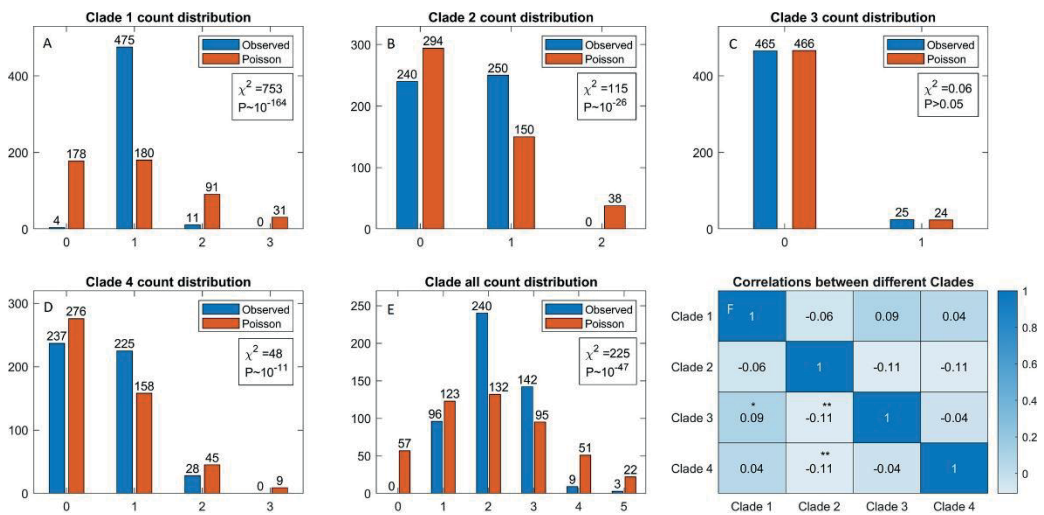


Fig. 2. Statistical analysis of GNAT-RHH TA pairs in *K. pneumoniae* strains. (A–D) Observed *vs* the Poisson distribution counts for clades 1–4 and (E) for all classes together. (F) Spearman's correlations between the counts of different GNAT-RHH clades in *K. pneumoniae* strains. P values of 0.05 and 0.1 are denoted by ** and *, respectively. Figure adapted from [2]

Conclusion: Consequently, based on our quantitative model we obtain that kacAT regulation results in monostable dynamics, and are able to explain all experimental data Moreover, we obtain that kacAT deletion does not lead to spontaneous persister formation, despite previous theoretical predictions, but in accordance with recent experimental evidence. This and our bioinformatics/statistical study challenges the assumption of cooperativity in TA action, possibly explaining the absence of spontaneous persister generation in kacAT. However, TA systems could be involved in antibiotic-induced persister formation, which provides an important future research direction.

Funding: This work is done in collaboration with Peifei Li, Ying-Xian Goh, Cui Tai, Zixin Deng, Zhaoyan Chen, Meng Wang, and Hui Wang. The study is supported by the Science Fund of the Republic of Serbia (grant No. 7750294, q-bioBDS). This work is also supported by the Science and Technology Commission of Shanghai Municipality (Grant No. 19430750600), the National Natural Science Foundation of China (Grant No. 32070572), the Medical Excellence Award Funded by the First Affiliated Hospital of Guangxi Medical University (No. XK2019025).

References

1. Li P., Goh Y.-X., Ilic B., Tai C., Deng Z., Chen Z., Djordjevic M., Ou H.-Y. Antibiotic-induced degradation of antitoxin enhances the transcription of acetyltransferase-type toxin-antitoxin operon. *J Antimicrob Chemother*. 2023;78(4):1066-1075. doi 10.1093/jac/dkad048
2. Goh Y.-X., Li P., Wang M., Djordjevic M., Tai C., Wang H., Deng Z., Chen Z., Ou H.-Y. Comparative Analysis of Diverse Acetyltransferase-Type Toxin-Antitoxin Loci in Klebsiella pneumoniae. *Microbiol Spectrum*. 2022;10(4):e0032022. doi 10.1128/spectrum.00320-22
3. Feng J., Kessler D.A., Ben-Jacob E., Levine H. Growth feedback as a basis for persister bistability. *Proc Natl Acad Sci USA*. 2014;111(1):544-549. doi 10.1073/pnas.1320396110

doi 10.18699/bgrs2024-2.2-15

Time series ordination for computational experiments dynamics (TSOCED) with biological models

Klimenko A.I.[1, 2] *, Vorobeva D.A.[3], Lashin S.A.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *klimenko@bionet.nsc.ru*

Key words: dynamic regimes; parametric stability; mathematical modelling; dynamic time warping; statistical analysis of computational experiments

Motivation and Aim: Computational biology widely uses mathematical models of biological systems, in particular, systems of ordinary differential equations, as well as models of dynamical systems described in other formalisms, including discrete models such as finite automata, Petri nets, Boolean networks, and individual-based models. Predicting changes in the type of solution dynamics as a function of changes in model parameters is a substantial scientific problem. However, this problem does not have analytical solution for most formalisms in a general case. Widely used practice of conducting a series of computational experiments, i. e. solving a series of direct problems with different sets of parameters followed by expert analysis of solution plots, becomes arduous with a large number of parameters and decreasing step of the parametric grid. Thus, developing analytical methods that are capable of working on a set of computational experiments in an aggregated form is of eminent importance.

Methods and Algorithms: The mathematical package Scilab was used to calculate the set of solutions of the studied models of biological systems and to generate a set of artificial dynamics. Time series analysis using dynamic time warping (DTW) algorithm and principal coordinates analysis (PCoA) was carried out in the R system using dtw and ape packages, respectively. A modified k nearest neighbours algorithm (weighted kNN) was used to construct a solution classifier.

Results and Conclusion: A method of visualisation and classification of different dynamic regimes of the model was developed using a composition of dynamic time warping algorithm and principal coordinate analysis. Time series ordination for computational experiments dynamics (TSOCED) visualises the results of multiple solutions of a mathematical model in the space of reduced dimensionality. TSOCED also allows to reveal a correspondence between the parameter values and the type of dynamic modes of the model's solutions. The use of the solution classifier by type of dynamics allows to carry out an automatic evaluation of the set of solutions obtained as a result of computational experiments with models aimed at studying parametric and initial data stability. This method was tested on the Lotka-Volterra model, artificial sets of different dynamics, and applied to the gene network model.

Funding: The study is supported by the Budget Project FWNR-2022-0020.

doi 10.18699/bgrs2024-2.2-16

Democratized Bioinformatics: Enabling accessibility through an integrated platform for drug discovery and multi-omics analysis

Mohammad T.[1, 2]*, Khan S.[2], Hassan M.I.[2]*

[1] *Department of Biophysics, All India Institute of Medical Sciences (AIIMS), New Delhi 110029, India*

[2] *Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi 110025, India*

** tajm@aiims.edu; mihassan@jmi.ac.in*

Keywords: Computational biology; Bioinformatics platform; Drug discovery; Multi-omics data analytics; Democratization of bioinformatics

Motivation and Aim: The rapidly evolving landscape of modern drug discovery and therapeutic management relies heavily on bioinformatics and computational biology tools. However, these resources often pose accessibility and affordability challenges, particularly for researchers with limited resources or expertise. To address these issues, there is a need for an integrated bioinformatics platform that combines state-of-the-art algorithms and AI models into a user-friendly interface, specifically tailored for modern drug discovery pipelines and multi-omics data analytics. This platform aims to democratize bioinformatics techniques, empower researchers across diverse scientific domains, and revolutionize the utilization of bioinformatics and computational biology.

Methods and Algorithms: The development of this platform involved a comprehensive methodology that includes requirement analysis, module development, data collection and preprocessing, module training and evaluation, platform development, testing and optimization, deployment and accessibility, validation and user feedback, dissemination and impact evaluation, and long-term maintenance and sustainability. Specialized modules focusing on various tasks such as virtual screening, molecular simulations, QSAR modeling, pharmacophore mapping, and NGS analytics have been developed and integrated into the platform using Python and R programming languages. Various AI models were trained and evaluated using diverse datasets to ensure accuracy and reliability.

Results: Several specialized programs have already been developed, deployed, and published [1, 2]. These programs perform tasks such as single-click high-throughput virtual screening and sequence and structure-based protein analyses. Currently, multiple programs/bioinformatics tools and algorithms are being developed and integrated into an integrated platform. The anticipated outcomes include the development of an integrated bioinformatics platform, specialized modules, benchmarked algorithms, an open-access platform, and case studies showcasing the platform's practical applications in drug discovery and multi-omics analyses. These outcomes aim to enhance research capability, democratize bioinformatics, contribute to national missions in biotechnology and healthcare, accelerate drug discovery, enable data-driven healthcare, foster scientific collaboration, have an educational impact, and increase publication and visibility.

Conclusion: This project addresses critical challenges in modern drug discovery and computational biology by developing an integrated bioinformatics platform that is accessible, user-friendly, and tailored for diverse research needs. By democratizing bioinformatics techniques and empowering researchers, this platform has the potential to accelerate scientific discoveries and foster collaborative research endeavors worldwide.

Acknowledgment: Taj Mohammad acknowledges SERB National Post Doctoral Fellowship (NPDF) with reference number: PDF/2022/001793.

References

1. Mohammad T., Mathur Y., Hassan M.I. InstaDock: A single-click graphical user interface for molecular docking-based virtual high-throughput screening. *Brief Bioinform*. 2021;22(4):bbaa279
2. Mathur Y., Mohammad T., Anjum F. et al. PyPAn: An automated graphical user interface for protein sequence and structure analyses. *Prot Pept Lett*. 2022;29(4):306-12

doi 10.18699/bgrs2024-2.2-17

Changes in the dynamics of a sex-structured population with non-overlapping generations during evolution

Revutskaya O.L.[1]*, Neverova G.P.[2], Zhdanova O.L.[2], Frisman E.Ya.[1]

[1] *Institute for Complex Analysis of Regional Problems, FEB RAS, Birobidzhan, Russia*

[2] *Institute of Automation and Control Processes, FEB RAS, Vladivostok, Russia*

* *oksana-rev@mail.ru*

Key words: population dynamics; sex structure; genetic structure; density-dependent regulation; dynamic modes; bistability; genetic variety

Motivation and Aim: The present work is devoted to the development of an eco-genetic approach to modeling the dynamics of limited sex-structured populations with non-overlapping generations within mathematical biology. We propose a dynamic model of a population that is structured by sex, which allows for studying evolutionary processes [1]. Note that adding sex structure to the model is important for understanding population abundance fluctuations, since sexual reproduction is one of the essential mechanisms of microevolutionary transformations and animal adaptation to the environment. This paper aims to study the evolutionary dynamics of an ecologically limited population that is structured by sex with seasonal reproduction and non-overlapping generations. The reason for choosing populations with non-overlapping generations and a short life cycle with a limited breeding period is the occurrence of significant fluctuations in the abundance of such species.

Methods and Algorithms: This paper proposes and studies a discrete-time model for a sex-structured population with non-overlapping generations. The model assumes that density-dependent factors limit the survival of females and males at different intensities. In addition, we consider a population with genetic diversity in individuals' reproductive potentials, controlled by one autosomal diallelic locus. We consider a panmictic population with Mendelian inheritance rules. To study the model we used the analytical and numerical methods of dynamic systems studies.

Results: This study showed that an increase in the average value of the reproductive potential in a population with density-dependent regulation of survival destabilizes the dynamics of abundance of females and males. In this case, the scenario of stability loss of equilibrium via the period doubling or Neimark–Sacker bifurcations depends on the intensity of selfregulation. The genetic composition of the population, namely, monomorphic genotype fixation or coexistence of all genotypes in polymorphism, is shown to be determined by the values of the reproductive potentials of heterozygotes and homozygotes, initial conditions, and a parameter describing the ratio of newborn females to males. In particular, we found that the increased reproductive potential of heterozygotes leads to a stable polymorphism within a population. The intermediate value of the reproductive potential of heterozygotes results in a monomorphism with an allele with greater fitness. The reduced reproductive potential of heterozygotes induces bistable dynamics when the different monomorphic fixed points with zero and non-zero population sizes coexist. Thus, depending on the values of the initial allele frequencies and the parameter describing the ratio of newborn females to males, the population either becomes extinct or survives by fixing one of the alleles during evolution. In this scenario, the

polymorphic state is unstable and is part of a transitional process in the dynamics to achieve one of the monomorphic states. With the coexistence of two monomorphic equilibria with zero and non-zero population sizes, the model of a population with two sexes proposed in this study reveals the possibility of bottleneck effects and evolutionary rescue. This model reveals two variants of evolutionary rescue: when the population recovers due to rare alleles with higher fitness or the emergence of new alleles caused by mutations.

Conclusion: The growth rate at which the population survives and develops is shown to depend on the fitness of the genotypes and the secondary sex ratio. As a result, the asymptotic genetic composition of the population is determined by the values of the reproductive potentials of the heterozygote and homozygotes, the initial conditions, and the parameter describing the ratio of newborn females to males. With disruptive selection, the influence of external factors changing the current genetic composition of a population can alter the direction of evolution and lead to the extinction of a successful developing population or a gradual population recovery due to evolutionary rescue after a noticeable decline in its abundance [1].

Funding: This work was carried out within the framework of the state targets of the Institute for Complex Analysis of Regional Problems of the Far Eastern Branch of the Russian Academy of Sciences.

References

1. Revutskaya O., Neverova G., Zhdanova O., Frisman E. The Evolutionary Dynamics of a Sex-Structured Population with Non-Overlapping Generations. *Mathematics*. 2023;11(24):4971. doi 10.3390/math11244971

2 Симпозиум «Системная компьютерная биология»
Symposium “Systems computational biology”

doi 10.18699/bgrs2024-2.3-01

Многофункциональный программный конвейер обработки данных HTS, полученных с использованием высокомультиплексных праймерных панелей

Гуков Б.*, Стеценко И., Мацвай А., Шипулин Г.

Федеральное государственное бюджетное учреждение «Центр стратегического планирования и управления медико-биологическими рисками здоровью» Федерального медико-биологического агентства, Москва, Россия

** bgukov@cspfmba.ru*

Ключевые слова: NGS; патоген; метагеномика; программа; вирус

Мотивация и цель: Вирусы считаются самой распространенной формой жизни на нашей планете. В настоящее время известно почти 13 миллионов геномных последовательностей вирусов, и это число с каждым годом растет экспоненциально. При этом считается, что большинство известных вирусов, вызывающих заболевания у человека, первоначально имели зоонозный характер и были переданы людям из популяций животных. Поэтому важность мониторинга природных резервуаров инфекций не нуждается в пояснении. Здесь обращают на себя особое внимание рукокрылые, которые на сегодняшний день считаются одним из самых важных природных резервуаров потенциально зоонозных инфекций, поскольку отличаются исключительным видовым разнообразием – они составляют примерно 20 % всех видов млекопитающих. Главное оружие в борьбе с будущими зоонозами – это превентивный контроль, мониторинг и оценка путей распространения известных патогенов, а также выявление новых патогенов в природных очагах инфекций. Выявление и изучение свойств новых, не описанных ранее вирусов тоже является важной задачей, поскольку достоверно неизвестно, какой именно патоген окажется способным вызвать следующую эпидемию зоонозного заболевания. Применение в данной области секвенирования, помимо выявления этиологического фактора заболевания, позволяет провести генотипирование любого локуса, реконструкцию полного генома для любого типа образца и любого типа патогена. Применение высокомультиплексных праймерных панелей широкого скрининга для таргетного обогащения позволяет повысить эффективность исследования, существенно снизив его себестоимость, в том числе за счет уменьшения количества данных секвенирования, необходимого для анализа одного образца. Однако, в силу частого использования при секвенировании «метода дробовика», существующие для решения поставленной задачи программы разработаны с расчетом на анализ именно таких образцов, что не позволяет произвести качественный и быстрый анализ данных другого типа. При этом за счет малого количества данных, необходимых для анализа каждого образца, количество образцов в одном запуске может быть значительно больше, чем при исследовании с использованием «методом дробовика», что приводит к тому, что время, необходимое на биоинформатическую обработку данных, начинает составлять значительную долю общего времени, необходимого для проведения исследования. Кроме того, существующие решения не специализируются на анализе образцов, полученных от животных, и

не позволяют провести обработку эндогенных примесей, в лучшем случае отфильтровывая их, без извлечения дополнительной информации.
Методы и алгоритмы: Программа pathogenid предназначена для анализа нуклеиновых кислот патогенов и предоставляет следующий функционал: фильтрация исходных прочтений по качеству и длине, объединение парных прочтений в один файл в случае парно-концевого секвенирования, обработка прочтений на предмет принадлежности к геному хозяина, кластеризация последовательностей с целью сокращения времени анализа и разделения объектов различных групп между собой, построение консенсусных последовательностей, сравнение схожести прочтений с последовательностями из референсных нуклеотидных баз данных, а в случае анализа нуклеиновых кислот вирусов и сравнение консенсусных последовательностей с базой данных аминокислотных последовательностей, аннотация прочтений, схожих с референсными последовательностями, по таксономической базе данных, расчет относительного и абсолютного содержаний нуклеиновых кислот, интересующих пользователя в контексте задачи, в исследуемом образце, создание отчета о результатах анализа образцов в форматах Excel-таблицы и JSON. Программа обрабатывает данные, поступающие с высокопроизводительных платформ секвенирования как второго, так и третьего поколений.
Результаты: Было разработано программное обеспечение pathogenid, адаптированное для задачи изучения вирома природных резервуаров инфекций. Программа апробирована на выборке из 713 биологических образцов, полученных от рукокрылых (Chiroptera), представителей 29 различных видов, в основном рода *Myotis* (N = 429, 60.17 %), пойманных по 16 субъектам Российской Федерации. Наибольшее количество образцов было собрано в Амурской (N = 165) и Алтайской (N = 111) областях. Исследование проводилось с использованием комбинации из семи праймерных панелей, состоящих суммарно из 40 415 первичных структур, нацеленных на обнаружение 52 родов вирусов, относящихся к 23 различным семействам. В результате анализа данных высокопроизводительного секвенирования были обнаружены последовательности нуклеиновых кислот 217 известных групп вирусов и 343 предположительно новых вирусов животных. Среди обнаруженных семейств чаще остальных встречалось семейство *Coronaviridae*. Представители семейства были обнаружены в 348 образцах, что составляет 48.1 % от всех проанализированных образцов. Полученные результаты, а также факты присутствия нуклеиновых кислот новых вариантов вирусов были подтверждены метагеномными исследованиями. Мы показали, что с использованием того же набора данных возможно эффективное определение таксономической принадлежности носителя вируса, что может быть важной задачей метагеномных исследований. Определение таксономии хоста возможно при условии наличия соответствующей геномной информации в открытых источниках. В среднем анализ одного образца, включая поиск гомологичных аминокислотных последовательностей, занимает 15 минут.
Выводы: В процессе работы была разработана программа pathogenid – универсальный биоинформатический конвейер обработки данных секвенирования, адаптированный для использования в задачах анализа вирома природных резервуаров инфекций. Данная программа подходит для образцов любой природы (при соответствующем методе экстракции), патогенов любой природы, обнаружения новых вирусов, анализа результатов секвенирования на любой платформе второго и третьего поколения. При наличии известных геномных данных животного-носителя,

pathogenid позволяет одновременно определять таксономическую принадлежность хоста. Пайплайн настроен для таксономической классификации патогенов с точностью до вида/штамма. Мы апробировали данный алгоритм анализа в проекте, включающим более 700 образцов, и подтвердили полученные результаты методом метагеномного секвенирования. Таким образом, мы можем проводить массовый скрининг биологического материала, обрабатывая большое количество образцов в короткие сроки, проводя при этом скрининг сразу на все значимые группы патогенов одновременно, выявляя в том числе новые, не известные ранее вирусы.

Финансирование: Исследование выполнено при поддержке государственного задания № 388-00084-24-00.

Multifunctional software pipeline for processing HTS data obtained using highly multiplexed primer panels

Gukov B.*, Stetsenko I., Matsvay A., Shipulin G.

Federal State Budgetary Institution "Centre for Strategic Planning and Management of Biomedical Health Risks" of the Federal medical and biological agency, Moscow, Russia

** bgukov@cspfmba.ru*

Key words: NGS; pathogen; metagenomics; programm, virsus

Motivation and Aim: Viruses are considered as the most prevalent form of life on our planet. Currently, almost 13 million genomic sequences of viruses are known and this number is growing exponentially every year. It is widely believed that the majority of known viruses responsible for human diseases originated from zoonotic transmission, meaning they were initially transmitted to humans from animal populations. Therefore, the significance of monitoring natural reservoirs of infections is self-evident. Bats (Chiroptera) merit particular attention in this regard, as they are now recognized as one of the most significant natural reservoirs of potentially zoonotic infections. Their exceptional species diversity, comprising roughly 20 % of all mammal species, underscores their importance in this context. The primary strategy in combating future zoonotic diseases involves preventive control, monitoring, and assessing the spread of both known pathogens and the identification of new pathogens in natural infection foci. Additionally, the identification and investigation of new, previously unknown viruses are crucial tasks, because pathogen, which will be capable of causing the next epidemic of zoonotic disease, cannot be accurately predicted. The utilization of sequencing goes beyond merely identifying the causative agent of the disease. It enables genotyping of any locus and reconstruction of the complete genome for any sample type and pathogen species. The use of broad screening panels for targeted enrichment can increase the efficiency of the study, significantly reducing its cost. The widespread adoption of the shotgun method in metagenomics has resulted in the specialization of existing software solutions for analyzing such samples. Consequently, high-quality and rapid analysis of other types of data is difficult. Additionally, current solutions lack specialization in analyzing animal samples and struggle to effectively manage endogenous contaminants, typically resorting to filtering without extracting additional information.

Methods and Algorithms: The "pathogenid" is pipeline intended for the analysis of nucleic acids of pathogens and provides the following functionality: filtering of source

reads by quality and length; combining paired reads into one file in the case of paired-end sequencing; processing of reads for belonging to the host genome; clustering of sequences in order to reduce analysis time and separate objects of different groups from each other; construction of consensus sequences; comparison of the similarity of reads with sequences from reference nucleotide databases; comparison of consensus sequences with a database of amino acid sequences, annotation of reads similar to reference sequences in a taxonomic database, calculation of the relative and absolute contents of nucleic acids of interest to the user in the context of the problem in the sample under study, creation of a report on the results of analysis of samples in Excel table and JSON formats. The program processes data coming from high-throughput second- and third-generation sequencing platforms.

Results: Pathogenid software has been developed and adapted for the task of studying the virome of natural reservoirs of infections. The program was tested using a set of 713 biological samples collected from Chiroptera, including representatives of 29 different species, mainly the genus *Myotis* (N = 429, 60.17 %), collected in 16 regions of the Russia. The largest number of samples were collected in the Amur (N = 165) and Altai (N = 111) regions. The study was conducted using a combination of seven primer panels, consisting of a total of 40,415 primary structures, aimed at detecting 52 genera of viruses belonging to 23 different families. Nucleic acids from 217 known groups of viruses and 343 presumably new animal viruses were discovered. The results obtained, as well as the facts of the presence of nucleic acids of new variants of viruses, were confirmed by metagenomics. Among all results, the *Coronaviridae* family was most common. Representatives of the *Coronaviridae* family were found in 348 samples, which is 48.1 % of all analyzed samples. We demonstrate that, utilizing the same dataset, it is possible to efficiently determine the taxonomic identity of the virus host, which can be an important goal of metagenomic studies. The determination of host taxonomy was achievable given the availability of suitable genomic information from open sources. On average, the analysis of one sample, inclusive of searching for homologous amino acid sequences, took 15 minutes.

Conclusion: Pathogenid – a universal bioinformatics pipeline for processing sequencing data, adapted for use in analyzing the virome of natural reservoirs of infections. This program is suitable for samples of any nature (with the appropriate extraction method), pathogens of any nature, detection of new viruses, analysis of sequencing results on any second and third generation platform. Given known genomic data from the host animal, pathogenid allows the taxonomic affiliation of the host to be simultaneously determined. The pipeline is configured for taxonomic classification of pathogens with species/strain accuracy. We tested this analysis algorithm in a project involving more than 700 samples and confirmed the results obtained using metagenomic sequencing. Thus, we are capable of conducting mass screening of biological material, processing a large number of samples swiftly. This enables simultaneous screening for all significant groups of pathogens, including the identification of new, previously unknown viruses.

Funding: The study was supported by the state assignment No. 388-00084-24-00.

doi 10.18699/bgrs2024-2.3-02

NeoCovasim: агентная эпидемиологическая модель с учетом транспортных потоков и ускорением на GPU

Манолов А.[1]*, Цуркис В.[2], Маслова И.[1], Козлов И.[1], Самойлов А.[1], Ильина Е.[1]
[1] *НИИ Системной биологии и медицины Роспотребнадзора, Москва, Россия*
[2] *ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)», Москва, Россия*
* *a.manolov@sysbiomed.ru*

Ключевые слова: агентное моделирование; вычислительная эпидемиология; COVID-19

Мотивация и цель: Недавняя пандемия COVID-19 ярко продемонстрировала недостаточность современных средств мониторинга, контроля и прогнозирования эпидемиологической ситуации в случае появления нового инфекционного агента. Это обусловливает необходимость дальнейшего развития подходов и методов, в том числе методов компьютерного моделирования, которые позволят лучше контролировать распространение эпидемии и при этом принимать сбалансированные и рациональные управленческие решения. Агентное моделирование активно применяется при симуляциях распространения эпидемий, поскольку позволяет учесть пространственную гетерогенность среды, индивидуальные особенности людей, демографические параметры и структуру контактов между людьми. При этом ограничивающим фактором их применения зачастую выступает большая вычислительная сложность данного подхода и трудность подбора параметров модели. Агентная модель Covasim стала популярной платформой для проведения агентного моделирования распространения SARS-CoV-2 [1]. К ее преимуществам можно отнести открытый исходный код, активную разработку, умеренный уровень сложности описания искусственной популяции и процесса передачи инфекционного агента. К ограничениям данной модели можно отнести невозможность моделирования нескольких населенных пунктов, связанных транспортными потоками, сравнительно большое время расчетов для крупных городов (с населением порядка миллиона человек и больше). Целью настоящей работы является разработка высокопроизводительной агентной модели, позволяющей моделировать популяцию, сравнимую по размеру и структуре с популяцией Российской Федерации и позволяющую учитывать региональные особенности и транспортные потоки между отдельными регионами.
Результаты: Нами был модифицирован код модели Covasim, что позволило моделировать множество отдельных населенных пунктов, которые связаны между собой транспортными потоками с заданной интенсивностью перемещения агентов. Вычисления по каждому отдельному населенному пункту проводятся параллельно, что позволяет эффективно использовать вычислительные ресурсы на многопроцессорных вычислительных кластерах. В графический интерфейс программы добавлена возможность независимо настраивать параметры отдельных населенных пунктов. Пользователь может загрузить файл с демографическими и статистическими показателями населенного пункта, на основе которых будет создана синтетическая популяция. Возможна настройка таких параметров, как половозрастная структура, распределение размера домохозяйств, школ и предприятий.

Также пользователь может задать интенсивность пассажиропотока между населенными пунктами. Вычислительные эксперименты выявили нелинейную связь между интенсивностью пассажиропотока и временной динамикой эпидемиологического процесса в населенных пунктах. Показано, что в случае связи транспортным потоком городов разного размера наблюдается сложная зависимость между размерами городов и временной динамикой эпидемического процесса. Эти результаты важны для интерпретации эмпирических данных по временной динамике количества инфицированных в различных населенных пунктах.
Для ускорения вычислений при моделировании крупных населенных пунктов нами реализовано ускорение расчетов за счет использования графического процессора (GPU). Было достигнуто 10–100-кратное ускорение (в зависимости от количества моделируемых агентов).
Нами была проведена оценка чувствительности модели к демографическим параметрам искусственной популяции на основе индексов Соболя. В качестве наблюдаемых значений использовалось суммарное количество инфицированных, количество умерших, пиковое количество инфицированных в день. Мы наблюдали, что наибольший вклад в эти значения вносят размеры домохозяйств, размеры классов в школах и размеры рабочих коллективов.
Выводы: На основе платформы с открытым исходным кодом Covasim нами была разработан программный комплекс NeoCovasim, позволяющий проводить параллельное моделирование нескольких населенных пунктов, связанных между собой транспортной сетью. Вычислительные эксперименты показали нелинейных характер связи между интенсивностью пассажиропотока и динамикой эпидемиологического процесса. Нами была реализована возможность проведения расчетов на GPU, что позволяет значительно ускорить процесс моделирования в случае крупных городов. Оценка чувствительности модели к параметрам показала, что наибольший вклад в поведение модели вносят размер домохозяйств, размер класса и размер рабочих коллективов.
Финансирование: Исследование поддержано субсидией Роспотребнадзора, номер 141-02-2023-208.

NeoCovasim: an agent-based epidemiological model with traffic flow and GPU acceleration

Manolov A.[1]*, Tsurkis V.[2], Maslova I.[1], Kozlov I. [1], Samoilov A. [1], Ilina E. [1]

[1] *Research Institute of System Biology and Medicine, Rospotrebnadzor, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, Moscow, Russia*

* *a.manolov@sysbiomed.ru*

Key words: agent-based modelling; computational epidemiology; COVID-19

Motivation and Purpose: The recent COVID-19 pandemic has clearly demonstrated the insufficiency of modern means of monitoring, controlling, and predicting epidemiological situations when faced with a new infectious agent. This necessitates the further development of approaches and methods, including computational modelling techniques, to better monitor the spread of epidemics and to make balanced and rational management decisions. Agent-based modelling is actively used in simulations of epidemic spread, as it allows for the consideration of spatial heterogeneity of the environment, individual

characteristics of people, demographic parameters, and the structure of contacts between individuals. However, the high computational complexity and the difficulty in selecting model parameters often limit its application.

The Covasim agent-based model has become a popular platform for modelling the spread of SARS-CoV-2 [1]. Its advantages include open-source code, active development, and a moderate level of complexity in describing the artificial population and the process of infectious agent transmission. However, limitations include the inability to model multiple settlements connected by traffic flows and relatively long calculation times for large cities (with populations of about or more than a million people).

The aim of this work is to develop a high-performance agent-based model that allows for the simulation of a population comparable in size and structure to the population of the Russian Federation, taking into account regional peculiarities and transport flows between individual regions.

Results: We modified the Covasim model to simulate a set of individual localities connected by traffic flows with a given intensity of agent movement. Computations for each settlement are performed in parallel, allowing efficient use of computing resources on multiprocessor clusters. The graphical interface of the program has been enhanced to allow independent configuration of parameters for individual settlements. Users can upload files with demographic and statistical indicators of a settlement, upon which a synthetic population is created. Parameters such as sex and age structure, household size distribution, schools, and enterprises can be configured. The intensity of passenger traffic between settlements can also be set.

Computational experiments revealed a non-linear relationship between the intensity of passenger traffic and the temporal dynamics of the epidemiological process in settlements. A complex dependence between city sizes and the epidemiological process dynamics was observed when cities of different sizes were linked by traffic flows. These results are crucial for interpreting empirical data on infection dynamics across different settlements.

To speed up calculations for large settlements, we implemented GPU acceleration, achieving a 10–100 times speed increase depending on the number of agents modelled.

We evaluated the sensitivity of the model to demographic parameters of the artificial population using Sobol indices. The total number of infections, the number of deaths, and the peak number of infections per day were the observed values. Household size, class size in schools, and workforce size were found to contribute most significantly to these values.

Conclusions: Based on the open-source Covasim platform, we developed NeoCovasim, a software package that allows parallel simulations of multiple localities connected by a transport network. Computational experiments showed a non-linear relationship between passenger traffic intensity and epidemiological dynamics. GPU acceleration significantly speed up the modelling process for large cities. Sensitivity analysis revealed that household size, class size, and workforce size are important parameters influencing model behavior.

Funding: The study is supported by the subsidy of the Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing No. 141-02-2023-208.

Список литературы/References

1. Kerr C.C., Stuart R.M., Mistry D., Abeysuriya R.G., Rosenfeld K., Hart G.R., Núñez R.C., Cohen J.A., Selvaraj P., Hagedorn B., George L. Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLoS Comput Biol.* 2021;17(7):e1009149

doi 10.18699/bgrs2024-2.3-03

The forecasting of the COVID-19 spread in Russian Federation regions based on conditional generative adversarial network

Krivorotko O.I.*, Zyatkov N.Yu.
Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
** o.i.krivorotko@math.nsc.ru*

Key words: deep learning; GAN; generative AI; machine learning; epidemiology; data processing

Motivation and Aim: COVID-19 caused by a novel coronavirus has continued to pose as a serious public health risk in 2024. The mutation of the virus has reached seasonal frequency, which requires additional resources of the medical system [1]. Despite global efforts to employ several health care strategies for minimizing the impact of the coronavirus on the community, there is still a great need to understand the dynamics of the virus as it transmits from human to human. Mathematical models with computational simulations have been very effective tools that help such global efforts to understand the dynamics of the disease, to estimate key transmission parameters and to make further improvements for controlling this disease [2-5]. The epidemic parameter estimation based on inverse problem approach using measurements about disease propagation such as diagnosed, severe, critical, vaccination and dead cases in fix time [3]. To forecast the spread of the COVID-19 epidemic, it is necessary to consider the socio-economic characteristics that influence the dynamics of the disease (full and/or partial lockdown, migration, emergence of a new strain, testing, vaccination, etc.). These characteristics affect changes in morbidity parameters, which makes the solution to the inverse problem non-unique and/or unstable to measurement errors.
Enough data on the dynamics of COVID-19 in the world since 2020 allows the use of statistical analysis and machine learning methods to build forecasts of epidemic time series. Classical machine learning and time-series approaches do not allow for probabilistic forecasts or have strong assumptions on the form of the future distribution of the target variable [6].
Methods and Algorithms: We apply Condition Generative Adversarial Networks (CGANs) [7] for probabilistic forecasting of the spread of COVID-19 in Saint-Petersburg using epidemic time series, such as new tested, diagnosed, hospitalized, critical (requiring a ventilator), immunity cases, deaths, self-isolation index in considered region [8] as well as new diagnosed COVID-19 cases in the World and its combinations [9] (architecture is presented in Fig. 1). CGAN is an unsupervised deep learning algorithm consists in two neural networks: the generator G generates new samples z of data close to the true data x and the discriminator D distinguishes generated samples of data from the true samples x with condition y. CGAN based on loss function

$$L(D,G) = E_{x\sim p_r(x)}[log\, D(x|y)\,] + E_{z\sim p_z(z)}\left[log\left(1 - D\big(G(y)\big)\right)\right].$$

We employ the ForGAN [10] architecture, which combines GANs with recurrent neural networks to generate probabilistic forecasts. During training, the generator tries to learn the distribution (under a given condition) of the original data. Then, after training the CGAN, the generator can build an arbitrarily large number of forecast trajectories (generate future data under a given condition from the learned distribution).

We prepare $M = 15$ time series epidemic data (above conditions y) for training and testing by the sliding time window method in the form of fix-target value pairs and divide them into training and test data. Sliding window width $L = 14$, forecast value $K = 5$. We use the previous 14-day data of COVID-19 in Saint-Petersburg to make a forecast for the next 5 days for the newly diagnosed cases, as well as the confidence interval of this forecast. The first step consists in data processing $f(t)$ based on logarithm transformation $f_{new}(t)=ln(f(t)+1)-ln(f(t-X)+1)$, $X=3,7,14$ days, interpolation, outlier removal and smoothing average [6] (see Fig. 2 for example).

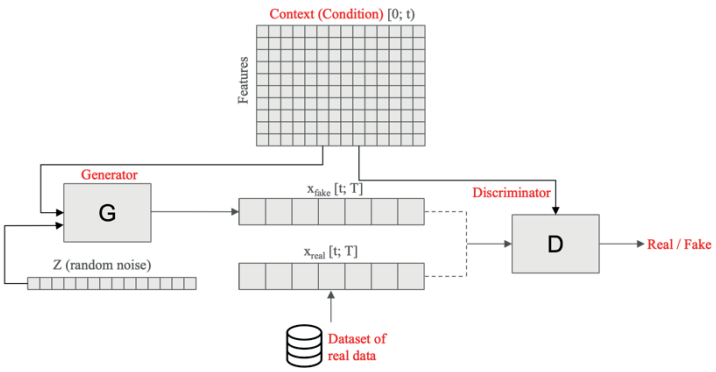


Fig. 1. Architecture of Conditional Generative Adversarial Network

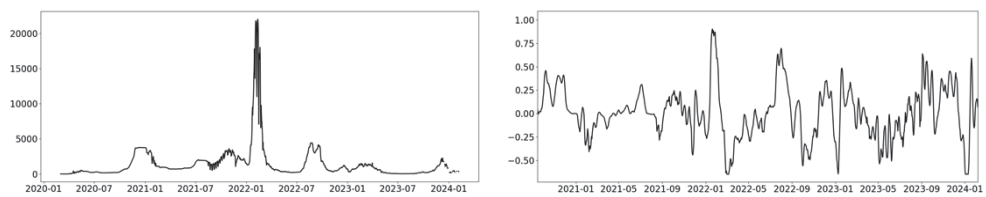


Fig. 2. Number of new diagnosed COVID-19 cases in Saint-Petersburg (left) and its 7-days logarithmic transformation (right)

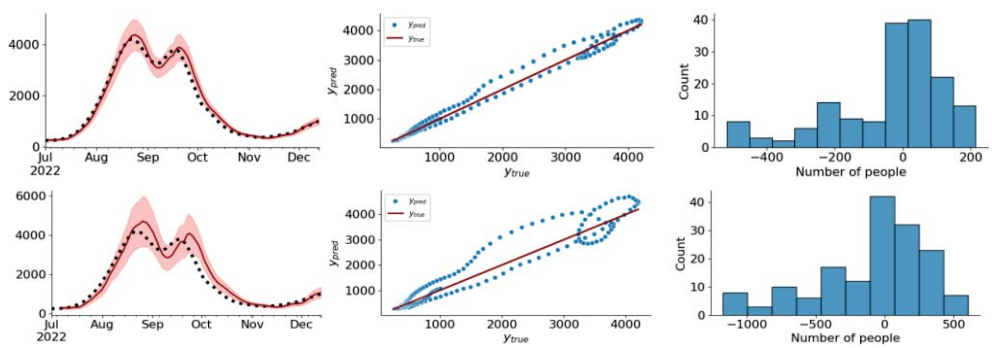


Fig. 3. First column: black dashed line – the real number of new COVID-19 diagnosed cases in St. Petersburg on day t+1 (the first row) and t+5 (the second row); light red area – confidence interval of the forecast on day t+1 and t+5, generated by G; the red line is the average of 10,000 generated forecasts. Second column: the red line is dependence of true forecasts on day t+1 and t+5 on themselves; blue dots – dependence of generator forecasts on true forecasts. Third column: histograms of absolute G forecast errors versus true forecasts

We will feed the discriminator with the target data dataset in order to train it to distinguish real forecasts from fake ones generated by the generator. For hints, we will also introduce a condition (data context) of size *LxM* in both *G* and *D*. *G* generates a prediction based on the white noise and the condition. *D* receives either a false or true prediction and a condition as input. Based on this, the *D* classifies whether the prediction is true or false.

Results: Predictions of newly diagnosed cases of COVID-19 in St. Petersburg were built for 1 and 5 days ahead based on CGAN and real data from 04.2020 to 02.2024 (Fig. 3). The percentage of successful hits in the confidence interval constructed by CGAN for 3-sigma rule for 1 day prediction is 75 % and for 5 days prediction is 67 %. The root mean square error is less than 15 %.

Conclusion: The prediction accuracy is slightly worse than that obtained from the full-connected neural network with and LSTM layers [6] and SEIR-HCD mathematical model based on the system of seven ordinary differential equations [11]. The amount of additional information for the GAN was less than when constructing a fully connected neural network and a SEIR-HCD model. The test data falls within the resulting confidence intervals constructed by the GAN. For more accurate forecasting, it is planned to optimize the network hyperparameters.

Funding: The study is performed to the Government research assignment for Sobolev Institute of Mathematics SB RAS (project FWNF-2024-0002).

References

1. Li Z., Zhang T. Analysis of a COVID-19 Epidemic Model with Seasonality. *Bull Math Biol*. 2022;84(12):146. doi 10.1007/s11538-022-01105-4
2. Krivorotko O.I., Kabanikhin S.I., Zyatkov N.Yu., Prikhodko A.Yu., Prokhoshin N.M., Shishlenin M.A. Mathematical modeling and forecasting of COVID-19 in Moscow and Novosibirsk region. *Num Anal Appl*. 2020;13(4):332-348. doi 10.1134/S1995423920040047
3. Krivorotko O., Kabanikhin S. Artificial intelligence for COVID-19 spread modeling. *J Inverse Ill-Posed Probl*. 2024;32(2). doi 10.1515/jiip-2024-0013
4. Krivorotko O., Sosnovskaia M., Vashchenko I., Kerr C., Lesnic D. Agent-based modeling of COVID-19 outbreaks for New York state and UK: parameter identification algorithm. *Infect Dis Modell*. 2022;7:30-44. doi 10.1016/j.idm.2021.11.004
5. Petrakova V., Krivorotko O. Mean field game for modeling of COVID-19 spread. *J Math Anal Appl*. 2022;514:126271. doi 10.1016/j.jmaa.2022.126271
6. Krivorotko O.I., Zyatkov N.Y., Kabanikhin S.I. Modeling Epidemics: Neural Network Based on Data and SIR-Model. *Comput Math Math Phys*. 2023;63(10):1929-1941. doi 10.1134/s096554252310007x
7. Wang S., Wang G., Fu Q., Song Y., Liu J. IH-TCGAN: Time-Series Conditional Generative Adversarial Network with Improved Hausdorff Distance for Synthesizing Intention Recognition Data. *Entropy*. 2023;25(5):781. doi 10.3390/e25050781
8. COVID-19 data for Saint-Petersburg: https://github.com/alexei-kouprianov/COVID.2019.ru
9. Vuletic M., Prenzel F., Cucuringu M. Fin-GAN: forecasting and classifying financial timeseries via generative adversarial networks. *Quant Finance*. 2024;24(2):175-199. doi 10.1080/14697688.2023.2299466
10. Koochali A., Schichtel P., Dengel A., Ahmed S. Probabilistic forecasting of sensory data with generative adversarial networks–ForGAN. *IEEE Access*. 2019;7:63868-63880. doi 10.1109/ACCESS.2019.2915544
11. Krivorotko O.I., Zyatkov N.Y. Data-driven regularization of inverse problem for SEIR-HCD model of COVID-19 propagation in Novosibirsk region. *Eurasian J Math Comput Appl*. 2022;10(1):51-68. doi 10.32523/2306-6172-2022-10-1-51-68

doi 10.18699/bgrs2024-2.3-04

Optimal control models for solving problems of mathematical epidemiology

Petrakova V.[1, 2]*, Krivorotko O.[2]

[1] *Institute of Computational Modelling, SB RAS, Krasnoyarsk, Russia*

[2] *Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia*

* *vika-svetlakova@yandex.ru*

Key words: epidemiological optimal control model; mean field games; compartmental models

Motivation and Aim: The Covid-19 pandemic has shown the need to develop strategies for regulating the policy by the planner (which, first of all, is government at various levels) of various social and economic measures to counter the virus. The practice of government decisions to curb the growth of morbidity, adopted in different countries, has shown that different strategies have different effects on the social and epidemiological situation (an example of various decisions and their consequences is discussed in [1–3]). Numerous examples are described in the social and epidemiological literature problems that arise when a social planner chooses a certain preventive strategy [4–6]. These include both social problems associated with distrust of information transmitted by health authorities and citizens' underestimation of the degree of risk, as well as economic restrictions associated with insufficient resources for carrying out quarantine measures. Thus, the problem of assessing scenarios for possible solutions for carrying out various antiviral measures in conditions of limited resources is relevant. So, the aim of this work is to show the result of many years of experience in using optimal control models in the field of mathematical epidemiology to solve problems of forecasting the spread of the virus, taking into account assumptions about the socio-economic characteristics of the region.

Methods and Algorithms: From the researcher's point of view, the solution to the above problems lies in the search for effective mathematical models that describe the situation, as well as methods for solving them. It is natural to assume that when solving problems of choosing optimal strategies, priority is given to mathematical models of optimal control. The report will consider two approaches to modeling the spread of epidemics developed by the authors: optimal control models with agent dynamics specified in the form of compartmental [7] epidemiological models, and the "mean field" models.

Mathematical "mean field" models, the original formulations of which were proposed in [8, 9], are becoming an increasingly popular tool for mathematical modeling. The advantage of the approach is the description of the collective behavior of a mass of agents (players) interacting in strategic situations using a small number of partial differential equations, which significantly reduces the calculation time and computational complexity of the problem compared to agent-based models when it comes to a large number of players. Mathematical epidemiology is no exception in this sense, since the task here is to describe the spread of a virus within a population represented by a large number of interacting individuals making strategic decisions (for example, about vaccination or compliance with specific quarantine restrictions). A broad overview of current work on the use of mean field models to simulate coronavirus dynamics is presented in [10].

In turn, optimal control models that depend only on time are not used so often to simulate the dynamics of the spread of an epidemic. This is apparently due to the fact that they differ structurally from compartmental epidemiological models only in the introduction of a control functional and a more complex set of numerical research methods. However, the practice of using such models shows a significant difference in forecasts depending on the type of model used. At the same time, simple optimal control models are more flexible in describing the system being modeled relative to the same "mean field" models, since the limitations imposed on them to justify the existence and uniqueness of the problem are less restrictive.

Results: The report will propose several formulations of the "mean field" problems and optimal control for the spread of the epidemic using the example of the Covid-19 pandemic in Novosibirsk and Krasnoyarsk with various assumptions about the social behavior of agents taking into account quarantine restrictions and the economic costs associated with the implementation of these restrictions. Their analytical and numerical comparison will be proposed.

Conclusion: Numerical comparison of optimal control models built on the basis of different approaches shows that "mean field" epidemiological models often give more accurate forecasts of the spread of morbidity compared to statistical data. Note, however, that "mean field" models are more sensitive to the choice of the problem functional compared to conventional optimal control models, since the domain of existence of the solution is significantly limited by the class of smooth Lipschitz functions, and the uniqueness of the solution can be guaranteed only in the case of strictly convex functionals and over short simulation periods. These limitations significantly complicate the selection of functional dependencies that describe the actual behavior of the economic-epidemiological system.

Funding: The study is supported by Russian Science Foundation, grant No. 23-71-10068.

References

1. Zanini D.S. et al. Practicing social isolation during a pandemic in Brazil: a description of psychosocial characteristics and traits of personality during covid-19 lockout. *Front Sociol*. 2021;6:615232. doi 10.3389/fsoc.2021.615232
2. Panda P.K. et al. COVID-19: lesson learnt from diagnostics to therapeutics. COVID-19: Lesson Learnt from Diagnostics to Therapeutics. In: Suar M., Misra N., Dash C. (Eds.). Microbial Engineering for Therapeutics. Singapore: Springer, 2022;345-374. doi 10.1007/978-981-19-3979-2_16
3. Chen J., Vullikanti A., Santos J., Venkatramanan S., Hoops S., Mortveit H., Lewis B., You W., Eubank S., Marathe M., Barrett C., Marathe A. Epidemiological and economic impact of COVID-19 in the US. *Sci Rep*. 2021;11(1):20451. doi 10.1038/s41598-021-99712-z
4. Gao X., Yu J. Public governance mechanism in the prevention and control of the COVID-19: information, decision-making and execution. *J Chin Governance*. 2020;5(2):178-197. doi 10.1080/23812346.2020.1744922
5. Coyne C.J., Duncan T.K., Hall A.R. The political economy of state responses to infectious disease. *South Econ J*. 2021;87(4):1119-1137. doi 10.1002/soej.12490
6. Capano G. Policy design and state capacity in the COVID-19 emergency in Italy: if you are not prepared for the (un) expected, you can be only what you already are. *Policy Soc*. 2020;39(3):326-344. doi 10.1080/14494035.2020.1783790
7. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics. *Proc R Soc London, Ser A*. 1927;115:700-721. doi 10.1098/rspa.1927.0118
8. Lasry J.-M., Lions P.-L. Mean field games. *Jpn J Math*. 2007;2(1):229-260. doi 10.1007/s11537-007-0657-8
9. Huang M., Malhame R.P., Caines P.E. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Commun Inf Syst*. 2006;6(3):221-251. doi 10.4310/CIS.2006.V6.N3.A5
10. Roy A., Singh C., Narahari Y. Recent advances in modeling and control of epidemics using a mean field approach. *Sådhanå*. 2023;48:207. doi 10.48550/arXiv.2208.14765

doi 10.18699/bgrs2024-2.3-06

Mean-field control models for describing information dissemination in online social networks

Zvonareva T.*, Krivorotko O.
Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia
** t.zvonareva@g.nsu.ru*

Key words: inverse problem, optimization, mean-field games, social process

Motivation and Aim: Mathematical models of the dynamics of social processes are poorly understood and rely on systems of differential equations [1, 2] and agent-based models [3]. Due to incomplete data, the use of these models does not fully characterize the process and its control. It is necessary to use the mean-field methodology [4] based on the dual game that allows one optimal control the information dissemination in online social networks. We consider a large number of users who can take states $x \in [0,1]$, where 0 means that the user is involved in the information dissemination process and 1 means the opposite. Then the density of users $u(x,t)$: $[0,1] \times [1,T] \to \mathbb{R}$ satisfies the Kolmogorov (Fokker-Planck) equation

$$u_t + \nabla(u\alpha) + \left(\frac{u}{K_{cap}} - 1\right) r(t)u - Du_{xx} = 0,\ x \in [0,1],\quad t \in [1,T] \qquad (1)$$

with initial condition and Robin boundary conditions

$$u(x,1) = \varphi(x),\ \ x \in [0,1] \qquad (2)$$

$$\beta_1 u_x(0,t) - \beta_2 u(0,t) = 0,\ \ u_x(1,t) = 0,\ \ t \in [1,T]. \qquad (3)$$

Here, $\alpha(x,t)$: $[0,1] \times [1,T] \to \mathbb{R}$ is the control parameter (in other words, the user strategy) that ensures the Nash equilibrium of the system of interacting agents and minimizes the cost functional with respect to (u,α)

$$J(u,\alpha) = \int_1^T \int_0^1 \left(d_1 e^{-t} \frac{u\alpha^2}{2} + d_2 (x-1)^2 u \right) dxdt.$$

Here d_1 and d_2 are the weight coefficients, K_{cap} is a carrying capacity, which is the maximum possible density of influenced users at a given distance, $r(t)$ is a growth rate of the number of influenced users placed at the same distance away from the source, D is a popularity of information which promotes the spread of the information through non-structure based activities such as search.

For $\alpha = 0$, equation (1) is investigated in [5].

Methods and Algorithms: Using the Lagrange multiplier method [6], a system similar to the Hamilton-Jacobi-Bellman equation is constructed:

$$\begin{gathered} v_t + \alpha v_x + \left(1 - \frac{2u}{K_{cap}}\right) r(t)v + Dv_{xx} = -d_1 e^{-t} \frac{\alpha^2}{2} - d_2 (x-1)^2, \\ v(x,T) = 0,\ \ x \in [0,1], \\ v_x(0,t) = v_x(1,t) = 0, \quad t \in [1,T]. \end{gathered} \qquad (4)$$

The optimality conditions are defined from following equations

$$d_1 e^{-t}\alpha + v_x = 0, \quad \alpha(0,t) = \alpha(1,t) = 0. \tag{5}$$

Problem (1)–(3), (4), (5) were solved by the finite-difference scheme proposed in [7]. To solve the inverse problem, it is necessary to determine the functions (φ, α) by additional measurements $u(x,t)$ in the computational domain. The identifiability of the model was analyzed by the Sobol sensitivity analysis method.

For the case $\alpha = 0$ (without control) we minimized the functional

$$J(\varphi) = \frac{(T-1)}{N_2} \sum_{k=1}^{N_2} \left| \sum_{i=1}^{N_1} u(x_i, t_k; \varphi) - f_k \right|^2 \tag{6}$$

using:

- the tensor optimization method (TT),
- the combination of TT and fast gradient method (FGM),
- the combined method with A.N. Tikhonov regularization,
- the combination of stochastic particle swarm optimization (PSO) and FGM.

Results: The following results were obtained for the case $\alpha = 0$ (without control). Figure 1 shows the results of recovering the initial density function $\varphi(x)$ at 6 points using: the tensor optimization method (TT) – red line with triangles, the combination of TT and fast gradient method (FGM) – green line with squares, the combined method with A.N. Tikhonov regularization – blue line with pentagons, and the combination of stochastic particle swarm optimization (PSO) and FGM – orange line with rhombuses.

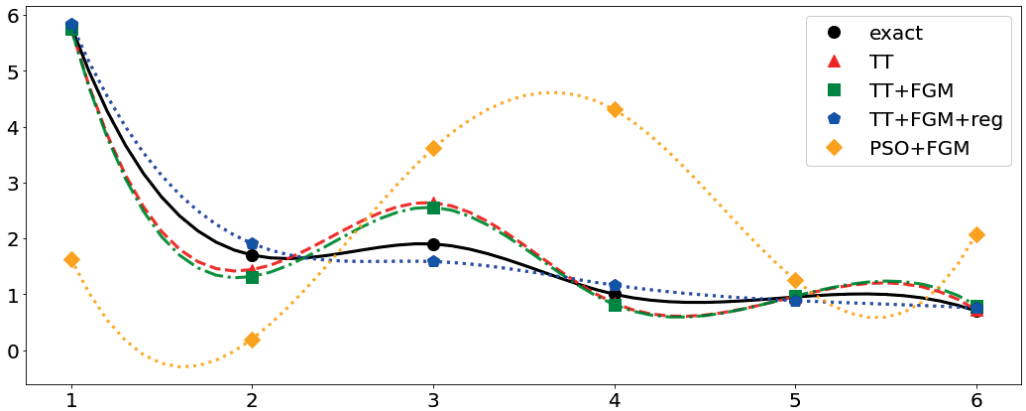
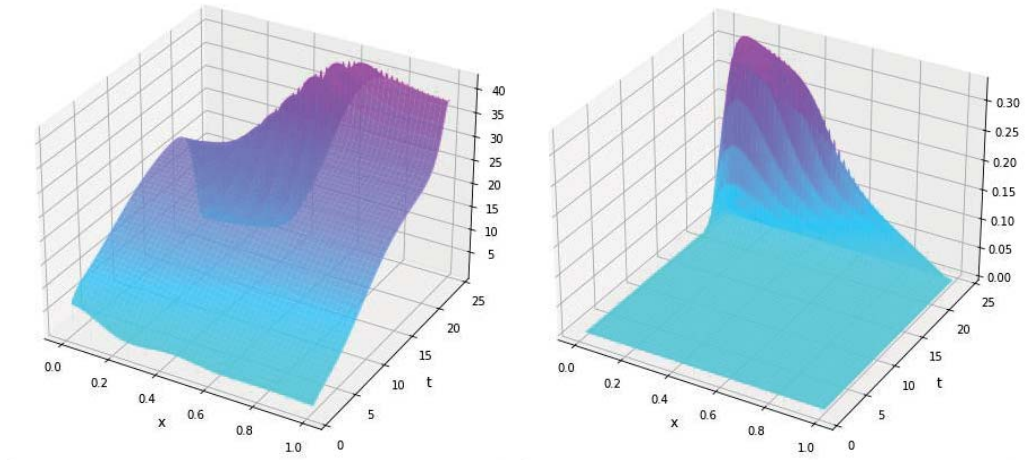


Fig. 1. Results of reconstruction of the initial density function $\varphi(x)$ for the case $\alpha = 0$ (without control)

The results of the numerical solution of the direct problem (determining the functions $u(x,t)$ and $\alpha(x,t)$ of problem (1)-(3), (4), (5) at a given $\varphi(x)$) were obtained. Figure 2 shows graphs (a) of the density of involved users $u(x,t)$ and (b) of the control function $\alpha(x,t)$.



(a) (b)

Fig. 2. Plots of (a) the density of involved users $u(x,t)$ and (b) the control function $\alpha(x,t)$

Conclusion: For the case $\alpha = 0$ (without control) the best solution of the inverse problem in terms of error value is the solution obtained by the regularized combination of TT and FGM [8].

The aim of the control process was to minimize the number of users involved in the process of information dissemination on the left boundary. It is shown that at the found control α the goal is achieved (the number of users involved in the process of information dissemination at x close to 0 is reduced) at the terminal moment of time.

Funding: The study is supported by the Russian Science Foundation (project No. 23-71-10068).

References

1. Goffman W., Newill V. Generalization of Epidemic Theory: An Application to the Transmission of Ideas. *Nature*. 1964;204:225-228
2. Wang F., Wang H., Xu K. Diffusive Logistic Model Towards Predicting Information Diffusion in Online Social Networks. In: Proc. 32nd Int. Conf. Distrib. Comput. Syst. Workshops. 2012;133-139
3. Goldenberg J., Libai B., Muller E. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Mark. Let.* 2001;12:211-223
4. Bauso D., Tembine H., Basar T. Opinion dynamics in social networks through mean-field games. *SIAM J Control Optim.* 2016;54(6):3225-3257
5. Zvonareva T.A., Krivorot'ko O.I. Comparative analysis of gradient methods for source identification in a diffusion-logistic model. *Comput Math Math Phys.* 2022;62(4):674-684
6. Bensoussan A., Frehse J., Yam P. Mean Field Games and Mean Field Type Control Theory. New York: Springer, 2013
7. Shaydurov V., Kornienko V. A finite-difference solution of mean field problem with a predefined control resource. *AIP Conf Proc.* 2020;2302:110004
8. Zvonareva T.A., Kabanikhin S.I., Krivorotko O.I. Numerical Algorithm for Source Determination in a Diffusion-Logistic Model from Integral Data Based on Tensor Optimization. *Comput Math Math Phys.* 2023;63(9):1654-1663.

3 Симпозиум «Структурная биология и фармакология: компьютерные и экспериментальные подходы»

Symposium “Structural biology and pharmacology: computational and experimental approaches”

doi 10.18699/bgrs2024-3.1-01

Анализ вторичной структуры нетранслируемых регионов РНК флавиподобных вирусов с сегментированным геномом

Алхиреенко Д.[1, 2], Гладышева А.[1]*

[1] *Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора, р.п. Кольцово, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *gladysheva_av@vector.nsc.ru*

Ключевые слова: РНК-содержащие вирусы; *Flaviviridae*; Jingmen tick virus; 5′–3′ UTR

Мотивация и цель: За последние пять лет с помощью высокопроизводительного секвенирования было обнаружено множество новых вирусов семейства *Flaviviridae*, обладающих атипичной структурой генома, что бросило вызов традиционным принципам классификации вирусов. Группа Jingmenvirus (JMV) состоит из новых РНК-содержащих вирусов, переносимых членистоногими, с неустановленной на данный момент патогенностью [1]. Kindia tick virus (KITV) – новый неклассифицированный флавиподобный вирус, относящийся к группе JMV семейства *Flaviviridae*, впервые обнаруженный в 2017 г. в иксодовых клещах на территории г. Киндия Гвинейской Республики [2, 3]. Он имеет сегментированный оцРНК(+) геном, состоящий из четырех сегментов, фланкированных 5′ и 3′ нетранслируемыми регионами (UTR). Два сегмента генетически и функционально связаны с генами, кодирующими неструктурные белки ортофлавивирусов, роль других двух сегментов не установлена. Этот вирус имеет необычное устройство генома, в котором структура и роль 5′ и 3′ UTR вирусной РНК в реализации его генетической информации на сегодняшний день фактически полностью остаются неизученными. Поэтому целью данной работы стало изучение вторичной структуры 5′ и 3′ UTR РНК сегментированных флавиподобных вирусов на примере нового вируса KITV.

Методы и алгоритмы: Новые неразрешенные мотивы 5′–3′ UTR (повторяющиеся шаблоны фиксированной длины) были найдены с использованием пакета программного обеспечения, предоставленного веб-сервером MEME Suite (версия 5.5.2). Прогнозирование структур 5′–3′ UTR и выравнивание последовательностей выполняли с использованием инструмента LocARNA. Полученные структурные выравнивания визуализировались как вторичные структуры РНК с помощью RNAplot из пакета ViennaRNA с использованием режимов «наиболее информативной последовательности» и «аннотирования ковариации пар оснований» и алгоритма построения графиков RNApuzzler. Вторичные структуры 5' и 3' UTR РНК KITV и их комплексов совместно со шпилькой MS2 были предсказаны с использованием следующих независимых инструментов: ViennaRNA Fold, UNA MFOLD 3.6 и RNAstructure. Моделирование проводили при температуре 37 °C и в ионных условиях 1 M NaCl без двухвалентных ионов. Параметры «максимальное расстояние между парными основаниями» (MDBPB) и «процент субоптимальности» (%S) подбирались вручную. Были установлены MDBPB от 60 до 100 и %S до 50 %. Верхняя граница количества вычисленных складок и верхняя граница общего количества одноцепочечных оснований,

которые разрешены в выпуклости или внутренней петле, были установлены на уровне 25. Остальные параметры были установлены по умолчанию, а начальная свободная энергия ΔG – на минимальное значение. Для детального анализа вторичные структуры были линеаризованы с помощью программы VARNA 3.9 и перерисованы в графических редакторах. Моделирование белковых структур производили в AlphaFold 2.

Сборку ДНК копий 5′ и 3′ UTR РНК KITV, их комплексов совместно со шпилькой MS2 и гена, кодирующего белок оболочки фага MS2, осуществляли методом сборочной ПЦР с использованием высокоточной ДНК полимеразы Q5 (NEB, США) и олигонуклеотидов длиной 60 оснований и последующим клонированием продуктов амплификации в плазмидный вектор pJET. Для создания экспрессирующей генетической конструкции использовали плазмидный вектор pEASY.

Результаты: В настоящей работе проведены моделирование и функциональная аннотация вторичных структур 5′ и 3′ UTR вирусов группы JMV и KITV, а также получены синтетические ДНК копии 5′ и 3′ UTR РНК KITV и произведен дизайн комплексов РНК-белок для последующей кристаллизации и РСА.

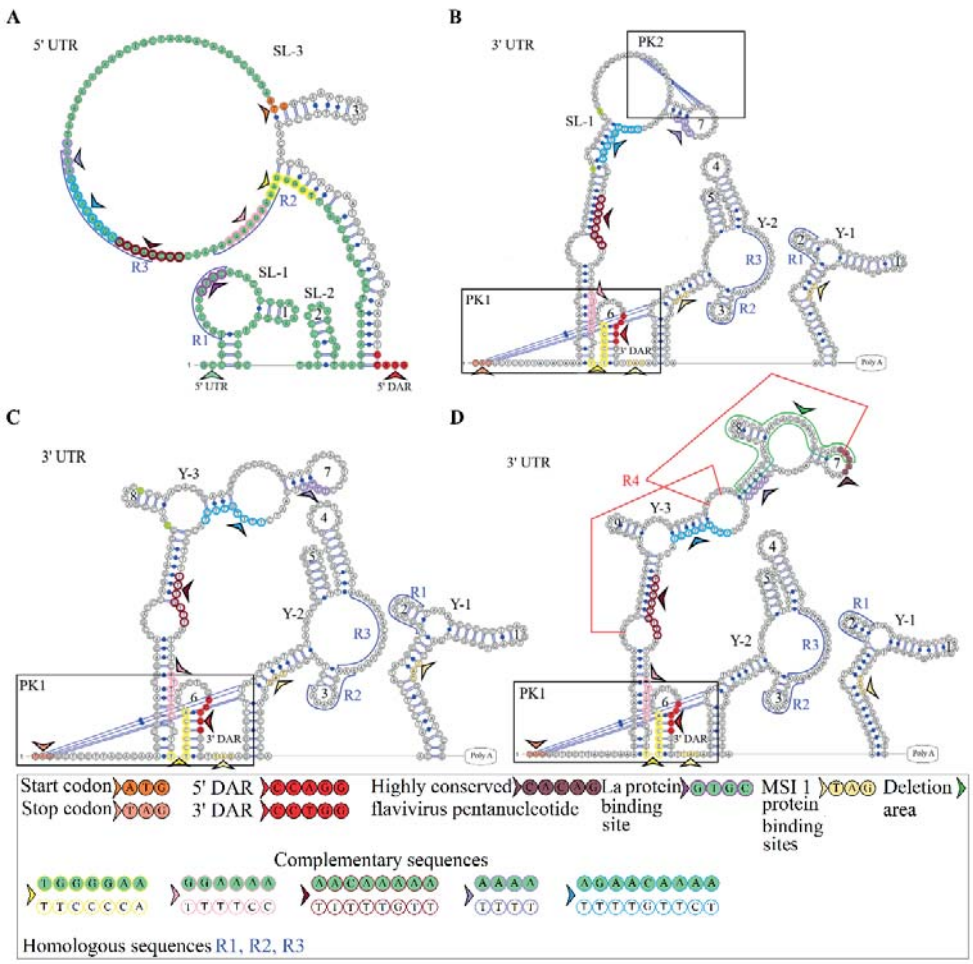


Рис. 1. Линейные модели вторичной структуры 5' UTR и 3' UTR сегмента 2 геномной РНК KITV. (A) 5' UTR изолята KITV/2018/1. (B) 3'UTR изолята KITV/2018/1. (C) 3' UTR изолята KITV/2018/2. (D) 3'UTR изолята KITV/2017/1

В ходе моделирования и анализа вторичных структур 5′ и 3′ UTR РНК KITV установлено, что для UTR различных сегментов KITV характерны: 1) полиаденилированные 3′ UTR; 2) мотивы 5′ DAR и 3′ DAR; 3) высококонсервативный пентануклеотид 5′-CACAG-3′; 4) регион-связывающий белок La; 5) множественные сайты UAG обеспечивающие взаимодействие с белком MSI1; 6) три гомологичные последовательности в 5′ UTR и 3′ UTR сегмента 2; 7) 3′ UTR сегмента 2 изолята KITV/2017/1 содержит два последовательных повтора по 40 нуклеотидов, образующих Y-3 структуру; 8) делеция в 35 нуклеотидов во втором повторе 3′ UTR 2-го сегмента изолятов KITV/2018/1 и KITV/2018/2, приводящая к модификации Y-3 структуры; 9) два псевдоузла PK1 и PK2 в 3′ UTR сегмента 2; 10) 5′ UTR и 3′ UTR представлены паттернами консервативных мотивов; 11) расположение последовательности 5′-AAGUG-3′ в первых шпильках UTR (рис. 1) [4]. Далее для получения экспериментальных данных о структуре 5′ и 3′ UTR были собраны полноразмерные ДНК копии 5′ и 3′ UTR РНК всех четырех сегментов KITV, содержащие T7 промоторную область для последующей транскрипции и область шпильки размером 18 нуклеотидов, взаимодействующей с белком оболочки фага MS2 для последующей кристаллизации комплексов РНК-белок. Также рекомбинантным путем в прокариотической системе клеток *E. coli* получен белок оболочки фага MS2, содержащий аффинную гексагистидиновую метку (6xHis-MS2). Моделированием структуры в AlphaFold 2 установлено, что химерный белок 6xHis-MS2 высокоструктурирован и аффинная метка не закрывает сайт связывания белка оболочки фага MS2 со специфичной 18-нуклеотидной шпилькой в составе 5′ и 3′ UTR РНК KITV.

Выводы: Таким образом, выявлены регуляторные элементы в структуре 5′ и 3′ UTR РНК KITV характерные для ортофлавивирусов, что позволяет предположить их функциональную значимость для репликации вирусов группы JMV и эволюционную схожесть ортофлавивирусов с флавиподобными сегментированными вирусами. Далее планируется кристаллизация полученных комплексов белка оболочки фага MS2 с 5′ и 3′ UTR РНК KITV и их рентгеноструктурный анализ. Это позволит экспериментально подтвердить структуру 5′ и 3′ UTR РНК KITV и установить роль нетранслируемых регионов вирусной РНК в реализации генетической информации новых флавиподобных вирусов с сегментированным геномом.

Финансирование: Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (соглашение от 12.10.2021 № 075-15-2021-1355) в рамках реализации отдельных мероприятий Федеральной научно-технической программы развития синхротронных и нейтронных исследований и исследовательской инфраструктуры на 2019–2027 гг.

Analysis of the secondary structure of the untranslated RNA regions of flavi-like viruses with a segmented genome

Alkhireenko D.[1, 2], Gladysheva A.[1]*

[1] *State Research Center of Virology and Biotechnology "Vector" Rospotrebnadzor, Koltsovo, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *gladysheva_av@vector.nsc.ru*

Key words: RNA viruses; *Flaviviridae*; Jingmenvirus; 5′–3′ UTR

Motivation and Aim: Over the past five years, high-throughput sequencing has discovered many new viruses in the family *Flaviviridae* that have atypical genome structures, challenging traditional principles of virus classification. The Jingmenvirus (JMV) group consists of novel arthropod-borne RNA viruses of currently unknown pathogenicity [1]. Kindia tick virus (KITV) is a new unclassified flavi-like virus belonging to the JMV group of the *Flaviviridae* family, first discovered in 2017 in ixodid ticks in the city of Kindia, Republic of Guinea. It has a segmented ssRNA(+) genome consisting of 4 segments flanked by 5' and 3' untranslated regions (UTR). Two segments are genetically and functionally related to genes encoding non-structural proteins of orthoflaviviruses; the role of the other two segments has not been established [2, 3]. This virus has an unusual genome structure, in which the structure and role of the 5′ and 3′ UTR of the viral RNA in the implementation of its genetic information remains virtually completely unexplored to date. Therefore, the goal of this work was to model and analyze the secondary structure of the 5′–3′ UTR RNA of segmented flavi-like viruses using KITV as an example.

Methods and Algorithms: Novel, unresolved 5′–3′ UTR motifs (recurring, fixed-length patterns) were searched using a software package provided by the MEME Suite web server (version: 5.5.2). Prediction of 5′–3′ UTR structures and that of sequence alignment were performed using the LocARNA tool. The resulting structural alignments were visualized as RNA secondary structures with RNAplot from the ViennaRNA package using the “most informative sequence” and “annotate covariance of base pairs” modes and the RNApuzzler plotting layout algorithm. Secondary structures of the KITV RNA 5′ and 3′ UTR and their complexes with MS2 coat protein were predicted using three independent tools as follows: ViennaRNA Fold, UNA MFOLD 3.6, and RNAstructure. Simulation was performed at a folding temperature of 37°C and under ionic conditions of 1 M NaCl without divalent ions. The parameters of “maximal distance between paired bases” (MDBPB) and “percent suboptimality” (%S) were selected manually. An MDBPB of 60 to 100 and a %S of up to 50 % were established. The upper bound on the number of computed folds and the upper bound on the total number of single-stranded bases that are allowed in a bulge or interior loop were set at 25. Other parameters were set as default, and the initial free energy ΔG was set as the minimum value. For detailed analysis, secondary structures were linearized using the VARNA 3.9 software and redrawn using graphic editors. Modeling of protein structures was carried out in AlphaFold2.

Assembly of DNA copies of the 5′–3′ UTR of KITV RNA, their complexes together with the MS2 hairpin and the gene encoding the MS2 coat protein was carried out by assembly PCR using high-precision Q5 DNA polymerase (NEB, USA) and oligonucleotides 60 bases long and subsequent cloning amplification products into the pJET plasmid. The pEASY plasmid was used to create an expression genetic construction.

Results: In this work, modeling and functional annotation of the 5' and 3' UTR secondary structures of the JMV group and KITV viruses were carried out, as well as synthetic DNA copies of the 5′–3′ UTR RNA KITV were obtained and RNA-protein complexes were designed for crystallization and X-ray diffraction analysis.

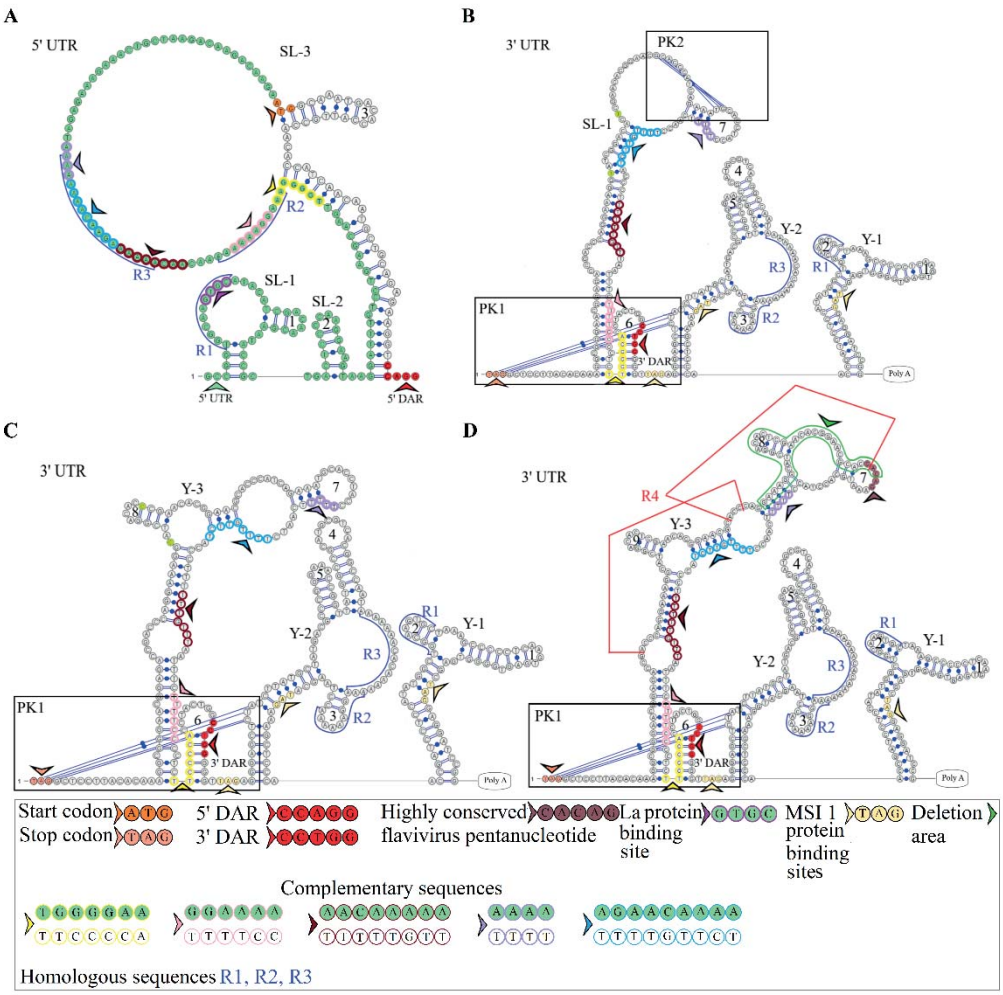


Fig. 1. Linear models of the secondary structure of 5′ UTR and 3′ UTR of genomic RNA segment 2 of KITV. (A) 5′ UTR of the KITV/2018/1 isolate. (B) 3′ UTR of the KITV/2018/1 isolate. (C) 3′ UTR of the KITV/2018/2 isolate. (D) 3′ UTR of the KITV/2017/1 isolate

UTRs of various KITV segments are characterized by the following points: (1) the polyadenylated 3′ UTR; (2) 5′ DAR and 3′ DAR motifs; (3) a highly conserved 5′-CACAG-3′ pentanucleotide; (4) a binding site of the La protein; (5) multiple UAG sites providing interactions with the MSI1 protein; (6) three homologous sequences in the 5′ UTR and 3′ UTR of segment 2; (7) the segment 2 3′ UTR of a KITV/2017/1 isolate, which comprises two consecutive 40 nucleotide repeats forming a Y-3 structure; (8) a 35-nucleotide deletion in the second repeat of the segment 2 3′ UTR of KITV/2018/1 and KITV/2018/2 isolates, leading to a modification of the Y-3 structure; (9) two pseudoknots in the segment 2 3′ UTR; (10) the 5′ UTR and 3′ UTR being represented by

patterns of conserved motifs; (11) the 5′-CAAGUG-3′ sequence occurring in early UTR hairpins (Fig. 1) [4].

Next, full-length DNA copies of the 5′–3′ UTR RNA of all four KITV segments, containing the T7 promoter region for subsequent transcription and an 18-nucleotide hairpin region that interacts with the MS2 coat protein for subsequent crystallization of RNA-protein complexes were constructed to obtain experimental data of the 5′–3′ UTR RNA KITV structure. Also, the MS2 phage envelope protein containing a hexahistidine affinity tag (6xHis-MS2) was obtained recombinantly in the *E. coli* cell. It was established that the chimeric protein 6xHis-MS2 is highly structured and the affinity tag does not overlap the binding site of the MS2 coat protein with a specific 18-nucleotide hairpin in the 5′–3′ UTR RNA KITV.

Conclusion: Thus, we identified regulatory elements in the UTRs of KITV, which are characteristic of orthoflaviviruses. This suggests that they hold functional significance for the replication of JMVs and the evolutionary similarity between orthoflaviviruses and segmented flavi-like viruses. Next, we plan to crystallize of complexes of the MS2 phage coat protein with the 5′ and 3′ UTR RNA KITV and their X-ray diffraction analysis. This will allow us to experimentally confirm the structure of the 5′ and 3′ UTR RNA KITV and establish the role of the viral RNA untranslated regions in the implementation of the genetic information of new flavi-like viruses with a segmented genome.

Funding: The study was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2021-1355 dated October 12, 2021) as a part of the implementation of certain activities of the Federal Scientific and Technical Program for the Development of Synchrotron and Neutron Research and Research Infrastructure for 2019–2027.

Список литературы/References

1. Zhang X., Wang N., Wang Z., Liu Q. The discovery of segmented flaviviruses: implications for viral emergence. *Curr Opin Virol*. 2020;40:11-18. doi 10.1016/j.coviro.2020.02.001
2. Gladysheva A.A., Gladysheva A.V., Ternovoi V.A., Loktev V.B. Structural Motifs and Spatial Structures of Helicase (NS3) and RNA-dependent RNA-polymerase (NS5) of a Flavi-like Kindia tick virus (unclassified Flaviviridae). *Vopr Virusol*. 2023;68(1):7-17. doi 10.36233/0507-4088-142 (in Russian)
3. Kartashov M.Y., Gladysheva A.V. et al. Molecular and genetic characteristics of the multicomponent flavi-like Kindia tick virus (Flaviviridae) found in ixodes ticks on the territory of the Republic of Guinea. *Vopr Virusol*. 2023;67(6):487-495. doi 10.36233/0507-4088-145 (in Russian)
4. Tsishevskaya A.A., Alkhireenko D.A., Bayandin R.B., Kartashov M.Y., Ternovoi V.A., Gladysheva A.V. Untranslated Regions of a Segmented Kindia Tick Virus Genome Are Highly Conserved and Contain Multiple Regulatory Elements for Viral Replication. *Microorganisms*. 2024;12(2):239. doi 10.3390/microorganisms12020239

doi 10.18699/bgrs2024-3.1-02

Использование нейросетевого алгоритма для предсказания аффинности связывания в белок-белковых комплексах ACE2-RBD

Богданова Е.[1]*, Чернухин А.[2], Шайтан К.[1], Новоселецкий В.[3]

[1] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

[2] *Российский химико-технологический университет им. Д.И. Менделеева, Москва, Россия*

[3] *Университет МГУ-ППИ в Шэньчжэне, биологический факультет, Шэньчжэнь, Китай*

* *eabogdanova.bioinf@gmail.com*

Ключевые слова: белок-белковые комплексы; аффинность связывания; коронавирус; нейронные сети

Мотивация и цель: Широкое распространение SARS-CoV-2 по сравнению с его предшественниками обусловлено более высокой аффинностью связывания S-белка с ACE2 [1]. Осознание данной закономерности повлекло за собой большое количество исследований интерфейса связывания RBD-ACE2, а также аминокислотных замен, влияющих на аффинность связывания между данными молекулами [2]. Физическое взаимодействие между молекулами белков имеет давнюю историю изучения с использованием многочисленных экспериментальных и вычислительных методов [3, 4], в том числе методов биоинформатики [5]. Одной из главных характеристик взаимодействия является константа диссоциации комплексов белок-белок (Kd). Создание алгоритмов аккуратной оценки этой энергии позволило бы более эффективно проводить направленный мутагенез взаимодействующих белков, что имеет существенное значение для создания медицинских препаратов белковой природы, включая антитела. В данной работе мы провели тестирование ряда ранее предложенных алгоритмов оценки энергии взаимодействия в комплексах «белок-белок» (Prodigy [6], FoldX [7], DFIRE [8], ROSETTADOCK [9]), а также разработанного нами алгоритма ProBAN [10].

Методы и алгоритмы: Для анализа аффинности связывания были отобраны комплексы RBD-ACE2 с известными пространственными структурами и значениями константы диссоциации (Kd). Итоговый набор данных содержит 48 комплексов ACE2 и RBD вирусов SARS CoV (3 структуры) и SARS CoV-2 (45 структур).

Предсказание аффинности связывания осуществлялось алгоритмом ProBAN, основанным на сверточной нейронной сети. Модель была обучена и протестирована на наборе белок-белковых комплексов с известными значениями константы диссоциации в базе данных PDBBind v.2020. Исследуемые комплексы RBD-ACE2 не содержатся в данной версии базы данных, что позволяет проводить предсказания на обученной модели без возможности «утечки» данных из обучающей выборки. Анализ полученных результатов, а также предсказание алгоритмом ProBAN осуществлялось с использованием Python дистрибутива Anaconda.

Результаты: Для анализа предсказания характеристик связывания в белок-белковых комплексах алгоритмом ProBAN было проведено его сравнение с предсказаниями, полученными веб-сервисом Prodigy. Данный метод осуществляет оценку аффинности связывания функцией, основанной на межмолекулярных контактах и признаках непосредственно на интерфейсе и полученных из анализа поверхности, не относящейся к интерфейсу взаимодействия. ProBAN (MAE = 0.5, корреляция Пирсона = 0.56) показывает более высокое качество предсказания по сравнению с Prodigy (MAE = 0.9, корреляция Пирсона = –0.38). Мы полагаем, что причиной этого является использование как более полной информации о взаимодействиях между атомами, так и большего порогового значения расстояния между атомами (10 Å), которое классифицирует пары атомов на взаимодействующие и нет. Используемое в Prodigy аналогичное пороговое значение расстояния между атомами (5.5 Å), по-видимому, отсеивает часть важных атомов, вносящих вклад в связывание.

Большое число алгоритмов, используемых для оценки аффинности связывания в белок-белковых комплексах, предсказывают не значение константы диссоциации, а свободную энергию Гиббса связывания. Для оценки работы этих алгоритмов (FoldX, DFIRE, ROSETTADOCK) на исследуемом наборе данных из полученных нами значений Kd мы рассчитали ΔG и сравнили с ΔG, полученной с использованием перечисленных алгоритмов. Рассчитанные значения метрик качества свидетельствуют о наибольшей эффективности ProBAN (MAE = 0.7 ккал/моль) среди всех проанализированных алгоритмов (для Prodigy 1.2 ккал/моль, у остальных выше).

Среди с комплексов с наибольшей абсолютной ошибкой (больше 1) подавляющая часть имела разрешение хуже 3 Å (6cs2, 7wk6, 7tex), и один имел разрешение 2.85 Å (8asy). Наблюдаемая закономерность свидетельствует о негативном вкладе нечетко разрешенного положения атомов в качество предсказания аффинности связывания, так как искажается информация об межатомных расстояниях, играющих ключевую роль во взаимодействии между белковыми молекулами.

Выводы: Собран набор данных, содержащий структуры комплексов ACE2-RBD с известными константами связывания, который может использоваться для изучения факторов, влияющих на связывание между RBD и ACE2, а также для обучения и тестирования алгоритмов, осуществляющих предсказание аффинности связывания. Использование нейросетевых методов, в частности ProBAN, для оценки константы диссоциации и свободной энергии Гиббса возможно не только между нативными RBD и ACE2, но и между другими мутантными формами. Полученные метрики качества свидетельствуют о недостаточной стабильности предсказаний для комплексов с высокой аффинностью связывания, что указывает на необходимость улучшения предикторов при увеличении набора доступных данных для анализа. Собранный набор данных, а таже результаты предсказания выложены в онлайн-репозитории, доступном по ссылке: https://github.com/EABogdanova/ProBAN_RBD-ACE2.

Финансирование: Исследование поддержано Российским научным фондом (№ 19–74–30003).

Using a neural network algorithm to predict binding affinity in ACE2-RBD protein-protein complexes

Bogdanova E.[1]*, Chernukhin A.[2], Shaitan K.[1], Novoseletsky V.[3]

[1] *Lomonosov Moscow State University, Moscow, Russia*

[2] *Mendeleev University of Chemical Technology of Russia, Moscow, Russia*

[3] *Shenzhen MSU-BIT University, Shenzhen, China*

* *eabogdanova.bioinf@gmail.com*

Key words: protein-protein complexes; binding affinity; coronavirus; neural networks

Motivation and Aim: The widespread distribution of SARS-CoV-2 compared to its predecessors was associated with the higher binding affinity of the S protein to ACE2 [1]. Awareness of this pattern led to many studies of the RBD-ACE2 binding interface, as well as amino acid substitutions that affect the binding affinity between these molecules [2]. The physical interaction between protein molecules has a long history of being studied using numerous experimental and computational methods [3, 4], including bioinformatics methods [5]. One of the main characteristics of the interaction is the dissociation constant of protein-protein complexes (Kd). The creation of algorithms for accurately assessing this energy would make it possible to carry out targeted mutagenesis of interacting proteins more effectively, which is essential for the creation of protein-based drugs, including antibodies. In this work, we tested several previously proposed algorithms for estimating the interaction energy in protein-protein complexes (Prodigy [6], FoldX [7], DFIRE [8], ROSETTADOCK [9]), as well as the ProBAN [10] algorithm we developed.

Methods and Algorithms: RBD-ACE2 complexes with known spatial structures and dissociation constant (Kd) values were selected for binding affinity analysis. The resulting data set contains 48 ACE2 and RBD complexes of the SARS-CoV (3 structures) and SARS-CoV-2 (45 structures) viruses.

Binding affinity prediction was performed by the ProBAN algorithm based on a convolutional neural network. This model was trained and tested on a set of protein-protein complexes with known dissociation constant values in the PDBBind v.2020 database. The RBD-ACE2 complexes under study are not contained in this version of the database, which allows predictions to be made on a trained model without the possibility of data "leakage" from the training set. Analysis of the results obtained, as well as prediction by the ProBAN algorithm, was carried out using the Anaconda Python distribution.

Results: To analyze the prediction of binding characteristics in protein-protein complexes by the ProBAN algorithm, it was compared with predictions obtained by the Prodigy web service. This method evaluates binding affinity by a function based on intermolecular contacts and features directly at the interface and obtained from analysis of surfaces not related to the interaction interface. ProBAN (MAE = 0.5, Pearson correlation = 0.56) shows superior prediction performance compared to Prodigy (MAE = 0.9, Pearson correlation = –0.38). We believe that the reason for this is the use of both more complete information about the interactions between atoms and a larger threshold value for the distance between atoms (10 Å), which classifies pairs of atoms

as those that interact or not. The similar interatomic distance threshold used in Prodigy (5.5 Å) appears to screen out some of the important atoms that contribute to binding.
Many algorithms used to estimate binding affinities in protein-protein complexes do not predict the value of the dissociation constant, but rather the Gibbs free energy of binding. To evaluate the performance of these algorithms (FoldX, DFIRE, ROSETTADOCK) on the data set under study, from the Kd values we obtained, we calculated ΔG and compared it with ΔG obtained using these algorithms. The calculated values of quality metrics for different algorithms indicate the highest efficiency of ProBAN (MAE = 0.7 kcal/mol) among all analyzed algorithms (for Prodigy 1.2 kcal/mol, for others higher).
Among the c complexes with the largest absolute error (more than 1), the overwhelming majority had a resolution worse than 3 Å (6cs2, 7wk6, 7tex) and one had a resolution of 2.85 Å (8asy). The observed pattern indicates a negative contribution of unclearly resolved atomic positions to the quality of binding affinity prediction, since information about interatomic distances, which play a key role in the interaction between protein molecules, is distorted.
Conclusions: A dataset containing the structures of ACE2-RBD complexes with known binding constants has been collected, which can be used to study factors influencing the binding between RBD and ACE2, as well as to train and test algorithms that predict binding affinities. The use of neural network methods, particularly ProBAN, to estimate the dissociation constant and Gibbs free energy is possible not only between native RBD and ACE2, but also other mutant forms. The obtained quality metrics indicate a lack of stability in predictions for complexes with high binding affinity, indicating the need to improve predictors as the data set available for analysis increases. The collected data set, as well as the prediction results, are posted in an online repository, available at the link: https://github.com/EABogdanova/ProBAN_RBD-ACE2.
Funding: The study is supported by the Russian Science Foundation (No. 19-74-30003).

Список литературы/References

1. Choudhary S., Sreenivasulu K., Mitra P. et al. Role of Genetic Variants and Gene Expression in the Susceptibility and Severity of COVID-19. *Ann Lab Med*. 2021;41(2):129-138
2. Vogel M., Augusto G., Chang X. et al. Molecular definition of severe acute respiratory syndrome coronavirus 2 receptor-binding domain. *Allergy*. 2022;**77**(1):143-149
3. Chothia C., Janin J. Principles of protein – protein recognition. *Nature*. 1975;256:705-708
4. Archakov I., Govorun V.M., Dubanov A.V. et al. Protein-protein interactions as a target for drugs in proteomics. *Proteomics*. 2003;3(4):380-391
5. Shi T.L., Li Y.X., Cai Y.D. et al. Computational Methods for Protein – Protein Interaction and Their Application. *Curr Protein Pept Sci*. 2005;6(5):443-449
6. Xue L.C., Rodrigues J.P., Kastritis P.L. et al. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics*. 2016;32(23):3676-3678
7. Schymkowitz J., Borg J., Stricher F. et al. The FoldX web server: an online force field. *Nucleic Acids Res*. 2005;33:W382-W388
8. Zhang C., Liu S., Zhou Y. Accurate and efficient loop selections by the DFIRE-based all-atom statistical potential. *Protein Sci*. 2004;13(2):391-399
9. Lyskov S., Gray J.J. The RosettaDock Server for Local Protein-Protein Docking. *Nucleic Acids Res.* 2008;36:W233-W238
10. Bogdanova E.A., Novoseletsky V.N. ProBAN: Neural network algorithm for predicting binding affinity in protein–protein complexes. *Proteins*. 2024. doi 10.1002/prot.26700

doi 10.18699/bgrs2024-3.1-03

Предсказание влияния замены отдельных аминокислотных остатков беровина, Ca^{2+}-регулируемого фотопротеина ктенофор, на условия его активации *in vitro*

Буракова Л.П.[1, 2]*, Иванисенко Н.В.[3, 1, 4], Рукосуева Н.В.[1], Иванисенко В.А.[3, 1], Высоцкий Е.С.[1]

[1] *Институт биофизики СО РАН, ФИЦ КНЦ СО РАН, Красноярск, Россия*

[2] *Сибирский федеральный университет, Красноярск, Россия*

[3] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[4] *AIRI, Москва, Россия*

* *burakoval@mail.ru*

Ключевые слова: биолюминесценция; фотопротеины; ктенофоры; беровин; молекулярное моделирование

Мотивация и цель: Биолюминесценцией, способностью излучать свет в видимом диапазоне спектра длин волн, обладают многочисленные наземные и морские живые организмы. Это явление наиболее распространено в морской среде. За свечение отвечают специальные белки, катализирующие окисление субстрата люциферина. Известно два основных их типа: люциферазы, работающие как классический фермент, и фотопротеины, образующие с субстратом устойчивый комплекс. Свечение гребневиков и гидромедуз обеспечивается Ca^{2+}-регулируемыми фотопротеинами, которые в бескальциевых условиях представляют собой комплекс апобелка с 2-гидропероксицелентеразином, а их свечение инициируется ионами кальция. Несмотря на схожие функции и пространственные структуры, фотопротеины гребневиков и гидромедуз разнятся по ряду свойств. Рекомбинантные фотопротеины ктенофор и медуз имеют низкую гомологию аминокислотных последовательностей и разные условия образования активного фотопротеина *in vitro*. Наиболее изученным фотопротеином ктенофор на сегодняшний день является беровин из *Beroe abyssicola*. Наибольший выход активного белка наблюдается при инкубации с целентеразином в щелочных условиях (pH 9.0) и при высокой концентрации соли (0.5 M NaCl) [1], тогда как для фотопротеинов гидромедуз это происходит при условиях, близких к физиологическим. Фотопротеины ктенофор обладают фоточувствительностью и инактивируются при облучении светом широкого диапазона длин волн [2], а для фотопротеинов гидромедуз такого эффекта не наблюдается. Также беровин является довольно термолабильным белком и теряет свою активность уже при 37 °C, что не характерно для белков медуз. Такие особенности фотопротеинов ктенофор накладывают ряд ограничений на их использование в качестве маркерных молекул при работе с эукариотическими клеточными линиями. Поскольку аминокислотный состав активного центра беровина кардинально отличается от такового гидромедуз, а пространственная структура его комплекса с 2-гидропероксицелентеразином не определена, то целью данной работы было предсказать аминокислотные замены в беровине, которые могут привести к сдвигу условий активации в область физиологических значений pH и концентрации соли.

Методы и алгоритмы: Вероятное положение аминокислотных остатков относительно субстрата в активном центре определяли из сравнения модели пространственной структуры беровина, полученной с помощью программы AlphaFold [3], с кристаллической структурой обелина в связанном с субстратом состоянии. С помощью нее в беровине был идентифицирован находящийся в гидрофобном окружении кластер полярных остатков, включая Lys90, который может образовывать сеть водородных связей с боковой цепью Asn107 и карбонильным кислородом основной цепи. Учитывая, что Lys90 положительно заряжен при нейтральном pH и нейтрален при щелочных pH, было сделано предположение, что его замена на алифатические аминокислоты может привести к изменению pH активации апофотопротеина. Для подтверждения этого с помощью сайт-направленного мутагенеза было получено 15 вариантов различных комбинаций аминокислотных замен с участием вышеуказанных аминокислотных остатков. Только 8 из них, K90A, K90M, N107S, K90A/N107S, K90M/N107S, K90A/W103F/N107S, K90M/W103F/N107S и K90M/W103F/N107V, показали способность к образованию фотопротеинового комплекса. Для каждого были определены условия активации относительно pH и концентрации NaCl. Все белки получены в высокоочищенном состоянии после инкубации апобелка с целентеразином и последующим отделением несвязавшихся компонентов с помощью ионообменной хроматографии, а их биолюминесцентные и спектральные свойства проанализированы.
Результаты: Для 6 мутантов беровина: K90A, K90M, K90A/N107S, K90M/N107S, K90A/W103F/N107S и K90M/W103F/N107S, были выявлены сдвиг условий активации в область физиологических значений pH и снижение зависимости от соли. Основной проблемой оказалось значительное снижение выхода активного белка при активации *in vitro* и удельной активности мутантов. Наибольший выход активного белка (~47 % от общего белка, полученного при хроматографии) в условиях активации при pH 6.0 и без NaCl показал K90M/W103F/N107S. При этом его удельная биолюминесцентная активность была почти в 3.5 раза ниже, чем у K90A, и составила 2.3 % от активности дикого беровина. Однако ее величина превышала таковую у K90M в 7.7 раза и практически совпадала с этим показателем для K90M/N107S. Но выход активного белка K90M/N107S был всего около 10 %. То есть добавление замены N107S к K90M увеличивает удельную активность, а добавление W103F повышает выход активного белка, как и было предсказано. Следует отметить, что подобного эффекта не наблюдалось для K90A, поскольку выход активного белка для K90A/N107S и K90A/W103F/N107S был одинаковым – около 10 %, а удельная активность только снижалась. Все 6 мутантов продемонстрировали снижение чувствительности к облучению светом по сравнению с диким беровином, а также некоторое повышение термостабильности. На спектр биолюминесценции в фотопротеинах влияют аминокислотные остатки, боковые радикалы которых находятся в непосредственной близости от 6-(p)-гидроксифенильной группы целентеразина. Согласно модели пространственной структуры беровина, Lys90 и Asn107 занимают как раз такую позицию. Максимумы спектров биолюминесценции всех 6 мутантов сдвинуты в фиолетовую область. Если спектр флуоресценции Ca^{2+}-разряженного беровина зависит от pH и показывает наличие разных ионных состояний целентерамида, то у всех исследованных белков этот показатель соответствовал только его нейтральной форме.

Выводы: Таким образом, экспериментальными данными подтверждена предсказанная с помощью методов молекулярного моделирования роль Lys90 и Asn107 в формировании условий активации беровина *in vitro* и получены мутанты с оптимумом активации при физиологических условиях и с повышенной фотостабильностью.
Финансирование: Исследование поддержано грантом РНФ «Рациональный дизайн целентеразин-зависимых биолюминесцентных белков с новыми свойствами с использованием методов мутагенеза и молекулярного моделирования» (№ 22-14-00125).

Prediction for the substitution of individual amino acid residues of berovin, the ctenophore Ca^{2+}-regulated photoprotein, to affect its activation conditions *in vitro*

Burakova L.P.[1, 2]*, Ivanisenko N.V.[3, 1, 4], Rukosueva N.V.[2], Ivanisenko V.A.[3, 1], Vysotski E.S.[1]

[1] *Institute of Biophysics, SB RAS, Krasnoyarsk, Russia*

[2] *Siberian Federal University, Krasnoyarsk, Russia*

[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[4] *AIRI, Moscow, Russia*

* *burakoval@mail.ru*

Key words: bioluminescence; photoproteins; ctenophores; berovin; molecular modeling

Motivation and Aim: Bioluminescence, i. e. the ability to emit light in the visible wavelength range, is characteristic of numerous marine and a few land organisms. This feature is attributed to the specific proteins that catalyze the oxidation of the substrate called luciferin. The two main types of such proteins are: luciferases, which act as classical enzymes, and photoproteins, which form a stable complex with the substrate. The luminescence of ctenophores and hydromedusae is initiated by calcium ions and is provided by Ca^{2+}-regulated photoproteins, which under calcium-free conditions are a complex of apoprotein with 2-hydroperoxycoelenterazine. Despite similar functions and spatial structures, the photoproteins of ctenophores and hydromedusae differ in a number of properties. The recombinant photoproteins of ctenophores and jellyfish have low amino acid sequence identity and different conditions for the formation of an active photoprotein *in vitro*. The now most-studied ctenophore photoprotein is berovin from *Beroe abyssicola*. Its highest yield is observed when it is incubated with coelenterazine under alkaline conditions (pH 9.0) and at high salt concentration (0.5 M NaCl) [1], while, e.g., for the hydromedusan photoproteins this occurs under conditions close to physiological ones. Ctenophore photoproteins are photosensitive and can be inactivated by irradiation with light of a wide range of wavelengths [2], whereas this effect is not observed in hydromedusan photoproteins. Berovin is also quite thermolabile and loses its activity already at 37 °C, which is not typical for jellyfish proteins. Such features of ctenophore photoproteins somewhat restrict their applicability as the marker molecules in eukaryotic cell lines. Since the amino acid composition of berovin active center is utterly different from that of hydromedusae, and the spatial structure of its complex with 2-hydroperoxycoelenterazine has not yet been determined, the goal of this work was to

predict the amino acid substitutions in berovin that can lead to a shift in activation conditions to the physiological pH values and salt concentration.

Methods and Algorithms: The probable position of amino acid residues relative to the substrate in the active site was determined by superimposing the model of the spatial structure of berovin, obtained by using the AlphaFold [3], onto the crystal structure of obelin in the substrate-bound conformation state. Using it, a cluster of polar residues, located in a hydrophobic environment, was identified in berovin, including Lys90, which can form a hydrogen network bonds with a side chain of Asn107 and the carbonyl oxygen of the main chain. Given that Lys90 is positively charged at neutral pH and is neutral at alkaline pH, it was hypothesized that its replacement with aliphatic amino acids could lead to a change in the pH of apophotoprotein activation. To confirm this, the 15 variants of various combinations of amino acid substitutions involving the above amino acid residues were obtained using site-directed mutagenesis. Only the 8 ones, K90A, K90M, N107S, K90A/N107S, K90M/N107S, K90A/W103F/N107S, K90M/W103F/N107S, K90M/W103F/N107V, showed the ability to form a photoprotein complex. For each of them, activation conditions were determined regarding pH and NaCl concentration. All proteins were obtained in a highly purified state after incubation of the apoprotein with coelenterazine and subsequent separation of the unbound components by using ion exchange chromatography; their bioluminescent and spectral properties were analyzed.

Results: For the 6 berovin mutants: K90A, K90M, K90A/N107S, K90M/N107S, K90A/W103F/N107S, and K90M/W103F/N107S a shift in activation conditions to the physiological pH values and a decrease in salt dependence were found. The main problem turned out to be a significant decrease in the yield of active protein upon activation *in vitro* and in the specific activity of the mutants. The highest yield of active protein (~47 % of its total amount obtained by chromatography) under activation conditions at pH 6.0 and without NaCl was shown for the K90M/W103F/N107S. Moreover, its specific bioluminescent activity was almost 3.5 times lower than that of K90A and amounted to 2.3 % of the activity of the wild berovin. However, its value exceeded that of K90M 7.7 times and practically coincided with the one of K90M/N107S. However the yield of the active K90M/N107S protein was only about 10 %. That is, adding the N107S to K90M increases the specific activity, and adding W103F increases the yield of the active protein, as was predicted. It should be noted that a similar effect was not observed for K90A, since the yield of the active protein for K90A/N107S and K90A/W103F/N107S was the same – about 10 %, and the specific activity only decreased. All the 6 mutants showed a decrease in sensitivity to light irradiation compared to the wild berovin, as well as a slight increase in thermostability.

The bioluminescence spectra of photoproteins are influenced by amino acid residues whose side radicals are closed to the 6-(p)-hydroxyphenyl group of coelenterazine. According to the model of the spatial structure of berovin, Lys90 and Asn107 occupy exactly this position. The maxima of the bioluminescence spectra of the mutants were shifted to the violet region. Whereas the fluorescence spectrum of Ca^{2+}-discharged berovin depends on pH and shows the presence of different ionic states of coelenteramide, this indicator corresponded only to its neutral form in all the proteins studied.

Conclusion: Thus, the role of Lys90 and Asn107 in the formation of conditions for the activation of berovin *in vitro* was experimentally predicted by molecular modeling

methods and the mutants with optimal activation under physiological conditions and with increased photostability were obtained.

Funding: The study was supported by RSF “Rational design of coelenterazine-dependent bioluminescent proteins with novel properties using mutagenesis and molecular modeling” (No. 22-14-00125).

Список литературы/References

1. Markova S.V., Burakova L.P., Golz S., Malikova N.P., Frank L.A., Vysotski E.S. The light-sensitive photoprotein berovin from the bioluminescent ctenophore *Beroe abyssicola*: a novel type of Ca(2+) – regulated photoprotein. *FEBS J*. 2012;279(5):856-870
2. Burakova L.P., Lyakhovich M.S., Mineev K.S., Petushkov V.N., Zagitova R.I., Tsarkova A.S., Kovalchuk S.I., Yampolsky I.V., Vysotski E.S., Kaskova Z.M. Unexpected Coelenterazine Degradation Products of *Beroe abyssicola* Photoprotein Photoinactivation. *Org Lett*. 2021;23(17):6846-6849
3. Jumper J., Evans R., Pritzel A., Green T., Figurnov M. et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589

doi 10.18699/bgrs2024-3.1-04

Исследование архитектуры светособирающих комплексов LH2 и LH1-RC из пурпурной серной бактерии *Ectothiorhodospira haloalkaliphila*

Бурцева А.Д.[1, 2]*, Баймухаметов Т.Н.[3], Чжан Х.[4, 5], Большаков М.А.[6], Ашихмин А.А.[6], Бойко К.М.[2]**

[1] *Московский физико-технический институт (Национальный исследовательский университет), Долгопрудный, Россия*

[2] *Институт биохимии им. А.Н. Баха, ФИЦ Биотехнологии РАН, Москва, Россия*

[3] *Национальный исследовательский центр «Курчатовский институт», Москва, Россия*

[4] *Шэньчжэньский институт передовых технологий, Китайская академия наук, Шэньчжэнь, Китай*

[5] *Школа наук о жизни, Южный университет науки и технологии, Шэньчжэнь, Китай*

[6] *Институт фундаментальных проблем биологии РАН, ФИЦ «Пущинский научный центр биологических исследований» РАН, Пущино, Россия*

** a.burtseva@fbras.ru; ** boiko_konstantin@inbi.ras.ru*

Ключевые слова: светособирающие комплексы; крио-ЭМ; фотосинтез; пурпурные серные бактерии; 3D структура

Мотивация и цель: Основными структурными компонентами фотосинтетического аппарата бактерий являются светособирающие комплексы (LH) и фотохимически активный реакционный центр (RC). Выделяют две группы светособирающих комплексов согласно их расположению относительно RC: прицентровые комплексы (LH1), которые образуют с RC так называемый коровый комплекс LH1-RC, и периферийные (LH2) комплексы. В отличие от несерных пурпурных бактерий, для серных бактерий существует только одна структура LH2 [1] и четыре LH1-RC [2–5] высокого и умеренного разрешения, что подчеркивает значительный недостаток структурных данных в этой области. Целью настоящей работы является использование метода крио-ЭМ для выявления структурных деталей комплексов LH1-RC и LH2 из пурпурной серной бактерии *E. haloalkaliphila*.

Методы и алгоритмы: 3 мкл образца LH2 нанесли на сетку QuantiFoil R0.6/1 Cu 300 mesh, которая не была обработана в тлеющем разряде перед нанесением образца. Сетка была витрифицирована погружением в жидкий этан с помощью прибора FEI Vitrobot Mark IV (Thermo Fisher Scientific, США). Крио-ЭМ данные собирали на просвечивающем электронном микроскопе Titan Krios 60-300 (Thermo Fisher Scientific, США), оснащенном детектором прямых электронов Gatan K3 Summit (Gatan, США) в режиме счета с помощью программы SerialEM 4.0 [6] в НИЦ «Курчатовский институт».

3 мкл образца LH1-RC нанесли на сетку QuantiFoil R0.6/1 Cu 300 mesh, которую гидрофилизировали с помощью системы очистки тлеющим разрядом PELCO easiGlow (Ted Pella. Inc) при 15 мА в течение 60 с. Сетка была витрифицирована погружением в жидкий этан с помощью прибора FEI Vitrobot Mark IV (Thermo Fisher Scientific, США). Сетки загрузили в электронный микроскоп Titan Krios с напряжением 300 кВ (Thermo Fisher), оснащенный детектором Gatan K3 Summit

(Gatan, США). Обработка данных проводилась в программе CryoSPARC ver. 4.2.1 [7]. Моделирование и уточнение выполнялись с помощью программы COOT [8].
Результаты: Методом крио-ЭМ получены структуры комплексов LH2 и LH1-RC из пурпурной серной бактерии *E. haloalkaliphila* с разрешением 1.7 и 2.5 Å соответственно. Комплекс LH1 имеет кольцевую архитектуру и состоит из 16 идентичных субъединиц с RC, расположенным в центре и состоящим из цитохрома c, а также полипептидных цепей, которые связывают различные кофакторы, включая бактериохлорофилл (БХл), феофитин, каротиноиды и ионы железа. LH2 представляет собой кольцеобразный октамер, каждая субъединица которого содержит α- и β-полипептиды, нековалентно связывающие три молекулы БХл и одну молекулу каротиноида. Последовательность полипептидов и тип каротиноидов в комплексе были однозначно определены благодаря разрешению, близкому к атомному. Проведен сравнительный анализ полученных структурных данных.
Выводы: Координаты LH2 из *E. haloalkaliphila* были депонированы в RCSB Protein Data Bank под кодом доступа 8Z4V. Крио-ЭМ карты комплекса LH2 депонированы в Electron Microscopy Data Bank под номером доступа EMD-39770.
Финансирование: Исследование поддержано грантом Российского научного фонда (№ 23-74-00062).

Structural studies of the LH2 and LH1-RC light-harvesting complexes from the purple sulfur bacterium *Ectothiorhodospira haloalkaliphila*

Burtseva A.D.[1, 2]*, Baymukhametov T.N.[3], Zhang H.[4, 5], Bolshakov M.A.[6], Ashikhmin A.A.[6], Boyko K.M.[2]**

[1] *Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russia*
[2] *Research Center of Biotechnology, RAS, Moscow, Russia*
[3] *National Research Center "Kurchatov Institute", Moscow, Russia*
[4] *Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China*
[5] *School of Life Sciences, Southern University of Science and Technology, Shenzhen, China*
[6] *Pushchino Scientific Center for Biological Research, RAS, Pushchino, Moscow region, Russia*
** a.burtseva@fbras.ru; ** boiko_konstantin@inbi.ras.ru*

Key words: light-harvesting complex; cryo-EM; photosynthesis; purple sulfur bacteria; 3D structure

Motivation and Aim: The main structural components of the photosynthetic apparatus of bacteria are light-harvesting complexes (LH) and photochemically active reaction center (RC). There are two types of LH complexes according to their location relative to the RC: the near-center complexes (LH1), which together with the RC form the core LH1-RC complex, and the peripheral complexes (LH2). In contrast to non-sulfur purple bacteria, there are only one LH2 [1] and four LH1-RC [2–5] high-to-moderate resolution structures for sulfur bacteria, highlighting the significant lack of structural data in this area. The aim of the present work was to use cryo-EM technique to reveal structural details of LH1-RC and LH2 complexes from purple sulfur bacteria *E. haloalkaliphila.*
Methods and Algorithms: 3 μl of LH2 sample was applied to the QuantiFoil R0.6/1 300 mesh Cu grid, which was not glow-discharged prior to sample application. The grid

was plunge-frozen in liquid ethane using a FEI Vitrobot Mark IV (Thermo Fisher Scientific, USA). Cryo-EM data were collected on a Titan Krios 60-300 transmission electron microscope (Thermo Fisher Scientific, USA) equipped with a Gatan K3 Summit direct electron detector (Gatan, USA) in counting mode using SerialEM 4.0 [6] at the National Research Centre "Kurchatov Institute".

3 μl of LH1-RC sample was applied to the QuantiFoil R0.6/1 300 mesh Cu grid, which was glow-discharged under 15 mA for 60 s using an easiGlow (Ted Pella. Inc). The grid was plunge-frozen in liquid ethane using a FEI Vitrobot Mark IV (Thermo Fisher Scientific, USA). Sample grids were loaded onto a 300 kV TitanKrios microscope (ThermoFisher) operated at 300 kV and equipped with a Gatan K3 Summit detector (Gatan, USA). Data processing was performed in CryoSPARC ver. 4.2.1 [7]. Modeling and refinement were made using COOT [8].

Results: Structures of LH2 and LH1-RC complexes from the purple sulfur bacterium *E. haloalkaliphila* were determined using cryo-EM technique at 1.7 Å and 2.5 Å resolution, respectively. The LH1 complex has a ring architecture and consists of 16 identical subunits with the RC at the very center, composed of cytochrome c and polypeptide chains that bind various cofactors including BChl, pheophytin, carotenoids, and iron ions. LH2 is a ring-shaped octamer, each subunit containing α- and β-polypeptides that noncovalently bind three BChl molecules and one carotenoid molecule. The sequence of the polypeptides and the type of carotenoids in the complex were unambiguously determined due to near-atomic resolution of this complex. Comparative analysis of the obtained structural data was done.

Conclusions: Atomic coordinates of the LH2 from *E. haloalkaliphila* are deposited in the RCSB Protein Data Bank under accession code 8Z4V. The cryo-EM maps of the LH2 complex were deposited in the Electron Microscopy Data Bank under accession number EMD-39770.

Funding: The study was supported the Russian Science Foundation grant (No. 23-74-00062).

Список литературы/References

1. Gardiner A.T., Naydenova K., Castro-Hartmann P., Nguyen-Phan T.C., Russo C.J., Sader K., Hunter C.N., Cogdell R.J., Qian P. The 2.4 Å cryo-EM structure of a heptameric light-harvesting 2 complex reveals two carotenoid energy transfer pathways. *Sci Adv*. 2021;7:eabe4650. doi 10.1126/sciadv.abe4650
2. Yu L.J., Suga M., Wang-Otomo Z.Y., Shen J.R. Structure of photosynthetic LH1-RC supercomplex at 1.9 Å resolution. *Nature*. 2018;556(7700):209-213. doi 10.1038/s41586-018-0002-9
3. Tani K., Kanno R., Makino Y. et al. Cryo-EM structure of a Ca^{2+}-bound photosynthetic LH1-RC complex containing multiple αβ-polypeptides. *Nat Commun*. 2020;11(1):4955. doi 10.1038/s41467-020-18748-3
4. Tani K., Kobayashi K., Hosogi N. et al. A Ca2+-binding motif underlies the unusual properties of certain photosynthetic bacterial core light-harvesting complexes. *J Biol Chem*. 2022;298(6):101967. doi 10.1016/j.jbc.2022.101967
5. Tani K., Kanno R., Harada A. et al. High-resolution structure and biochemical properties of the LH1-RC photocomplex from the model purple sulfur bacterium, Allochromatium vinosum. *Commun Biol*. 2024;7(1):176. doi 10.1038/s42003-024-05863-w
6. Mastronarde D.N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol*. 2005;152(1):36-51. doi 10.1016/j.jsb.2005.07.007
7. Punjani A., Rubinstein J.L., Fleet D.J., Brubaker M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*. 2017;14(3):290-296. doi 10.1038/nmeth.4169
8. Emsley P., Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr, Sect D: Biol Crystallogr*. 2004;60(Pt. 12. Pt. 1):2126-2132. doi 10.1107/S0907444904019158

doi 10.18699/bgrs2024-3.1-05

Изучение механизмов активации рецептора инсулина (IR) с использованием мутантных форм рецептора

Гавриленкова А.А.[1, 2]*, Деев И.Е.[2], Бочаров Э.В.[1, 2], Серова О.В.[2]

[1] *Московский физико-технический институт (национальный исследовательский университет), Долгопрудный, Россия*

[2] *Институт биоорганической химии им. ак. М.М. Шемякина и Ю.А. Овчинникова РАН, Москва, Россия*

* *gavrilenkova.aa@phystech.edu*

Ключевые слова: тирозинкиназа; рецептор инсулина; трансмембранный домен; мутации

Мотивация и цель: Рецепторные тирозинкиназы (РТК) – очень важное звено при передаче сигнального каскада внутрь клетки. Они играют ключевую роль в развитии, дифференцировке, пролиферации и миграции клеток. РТК состоят из трех частей: внеклеточная часть, которая отвечает за связь с лигандом, трансмембранный (ТМ) домен и внутриклеточная часть, которая отвечает за дальнейшее фосфорилирование субстратов [1]. Семейство рецептора инсулина относится к РТК, и на данный момент точные механизмы их конформационных изменений при активации до сих пор выясняются. Предполагается, что в неактивном состоянии SS-связанного димера рецептора инсулина ТМ-домены находятся в конформации, препятствующей взаимодействию цитоплазматических частей молекулы. При связывании лиганда конформация димера рецептора меняется, внутриклеточные тирозинкиназные домены сближаются и фосфорилируют друг друга, вызывая клеточный ответ [2].

Методы и алгоритмы: Для изучения механизмов активации рецептора инсулина (IR) нами были получены мутантные формы рецептора, содержащие двойные и одиночные замены в ТМ-домене. Клетки HEK293 трансфицировали плазмидными конструкциями, кодирующими мутантные формы IR с заменами: I951E-F952R, F956E-S957R, I960E-G961R, I951E и F952R. Далее клетки инкубировали в среде F12 с добавлением инсулина. Клеточные лизаты анализировали методом вестерн-блота с антителами к общей и фосфорилированной форме IR.

Результаты: Мутации F956E-S957R и I960E-G961R не влияли на характер активации рецептора, в то время как замены I951E-F952R, I951E и F952R приводили к фосфорилированию рецептора в отсутствие лиганда, в отличие от рецептора дикого типа, который фосфорилируется только в присутствии инсулина. Мы предполагаем, что данные замены приводят к стабилизации димера рецептора в активном состоянии. Нами было установлено, что автофосфорилирование рецептора IR с заменами I951E-F952R, I951E и F952R в отсутствие инсулина приводит к активации внутриклеточных сигнальных белков IRS-1 и ERK. Это свидетельствует о том, что данные замены в трансмембранном домене IR, вероятно, приводят к образованию функционально активного димера рецептора в отсутствие лиганда.

Выводы: Нами показано, что ТМ-домен играет важную роль в активации рецептора инсулина. Даже точечные замены в его аминокислотной последовательности могут приводить к получению функционально активных

форм рецептора, предположительно, за счет образования водородных связей и солевых мостиков между ТМ-доменами внутри липидного бислоя.
Финансирование: Исследование выполнено за счет гранта Российского научного фонда (№ 23-74-00024).

Study of the mechanisms of activation of the insulin receptor (IR) using mutant forms of the receptor

Gavrilenkova A.A.[1, 2]*, Deyv I.E.[2], Bocharov E.V.[1, 2], Serova O.V.[2]

[1] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

[2] *Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

* *gavrilenkova.aa@phystech.edu*

Key words: tyrosine kinase; insulin receptor; transmembrane domain; mutations

Motivation and Aim: Receptor tyrosine kinases (RTKs) are a very important link in the transduction of the signaling cascade into the cell, they play a key role in the development, differentiation, proliferation and migration of cells. RTKs consist of 3 parts: the extracellular part, which is responsible for ligand binding, the transmembrane (TM) domain, and the intracellular part, which is responsible for further phosphorylation of substrates [1]. The insulin receptor (IR) family belongs to RTKs, and at the moment the exact mechanisms of their conformational changes during activation are still being investigated. It is assumed that in the inactive state of the TM-domains of the SS-bonded IR dimer are in a conformation that prevents the interaction of cytoplasmic parts of the receptor. Upon ligand binding, the conformation of the receptor dimer changes, intracellular tyrosine kinase domains converge and phosphorylate each other, causing a cellular response [2].
Methods and Algorithms: To study the mechanisms of the IR activation, we obtained mutant forms of the receptor containing double and single substitutions in its TM-domain. HEK293 cells were transfected with plasmid constructs encoding mutant forms of IR with the substitutions: I951E-F952R, F956E-S957R, I960E-G961R, I951E and F952R. Next, the cells were incubated in F12 medium with the addition of insulin. The cell lysates were analyzed by the Western blot method with antibodies to the total and phosphorylated forms of IR.
Results: As a result, mutations F956E-S957R and I960E-G961R did not affect the fashion of the receptor activation. Substitutions I951E-F952R, I951E and F952R resulted in phosphorylation of the receptor in the absence of a ligand, unlike the wild-type receptor, which is phosphorylated only in the presence of insulin. We assume that these substitutions lead to stabilization of the receptor dimer in the active state. We found that autophosphorylation of IR with substitutions I951E-F952R, I951E and F952R in the absence of insulin leads to activation of intracellular signaling proteins IRS-1 and ERK. This indicates that these substitutions in the IR TM-domain probably lead to the formation of a functionally active receptor dimer in the absence of a ligand.
Conclusion: We have shown that the TM-domain plays an important role in the activation of IR, and even point substitutions in its amino acid sequence can lead to the production of the functionally active forms of the receptor, presumably due to the formation of hydrogen bonds and salt bridges between the TM-domains within the lipid bilayer.

Funding: The study was supported by the Russian Science Foundation grant (No. 23-74-00024)

Список литературы/References

1. Lemmon M.A., Schlessinger J. Cell signaling by receptor-tyrosine kinases. *Cell.* 2010;141(7):1117-1134. doi 10.1016/j.cell.2010.06.011
2. Kuznetsov A.S., Zamaletdinov M.F., Bershatsky Y.V., Urban A.S., Bocharova O.V., Bennasroune A., Maurice P., Bocharov E.V., Efremov R.G. Dimeric states of transmembrane domains of insulin and IGF-1R receptors: Structures and possible role in activation. *Biochim Biophys Acta Biomembr*. 2020;1862(11):183417. doi 10.1016/j.bbamem.2020.183417

doi 10.18699/bgrs2024-3.1-06

Структурные особенности нуклеопротеинов новых ортонайровирусов: Yezo virus, Songling virus и Beiji nairovirus

Гладышева А.*, Тюрин В., Цишевская А., Иматдинов И., Агафонов А.
Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора, р.п. Кольцово, Россия
** gladysheva_av@vector.nsc.ru*

Ключевые слова: РНК-содержащие вирусы; *Orthonairovirus*; вирусные белки; нуклеопротеин

Мотивация и цель: Вирусы являются причиной серьезных заболеваний во всем мире, особенно в развивающихся странах. В связи с ростом численности, плотности и мобильности населения увеличивается риск возникновения вспышек известных и неизвестных вирусных инфекций. В последние годы с использованием технологии NGS было выявлено множество новых патогенных для человека вирусов, переносимых клещами, в том числе ортонайровирусов: Songling virus (SGLV) [1], Yezo virus (YEZV) [2, 3] и Beiji nairovirus (BJNV) [4]. Имеющиеся ограниченные данные об этих вирусах указывают на то, что с высокой долей вероятности они могут быть возбудителями лихорадки у человека, что представляет новую угрозу для общественного здравоохранения. При этом информация о структурных особенностях вириона и белков новых ортонайровирусов, являющаяся необходимой для рационального создания эффективных вакцин и терапевтических противовирусных препаратов, на данный момент полностью отсутствует. Поэтому целью нашей работы стало изучение структурных особенностей нуклеопротеинов (NP) SGLV, YEZV и BJNV, которые играют важную роль в жизненном цикле ортонайровирусов и содержат иммунодоминантные регионы.
Методы и алгоритмы: Полноразмерные последовательности сегмента 1 SGLV, YEZV и BJNV были взяты из базы данных GenBank. Поиск близкородственных нуклеопротеинов, пространственные структуры которых уже были решены экспериментально, выполнен с помощью NCBI BLAST (National Center for Biotechnology Information Basic Local Alignment Search Tool) по базе данных PDB с использованием алгоритма blastp (protein-protein BLAST). Множественное выравнивание аминокислотных последовательностей выполнено в программе Unipro UGENE с использованием итеративного метода выравнивания MUSCLE. Функциональная аннотация вирусных белков произведена с помощью программы Protomenal и веб-сервиса HMMER. Поиск мотивов проводился по базам данных Pfam и NCBI CDD (Conserved Domains Database). Моделирование пространственных структур NP SGLV, NP YEZV и NP BJNV производили на сервере AlphaFold. Отбор пространственных моделей для дальнейшего теоретического анализа производился на основе коэффициента достоверности для каждой аминокислоты с учетом прогнозируемой локальной разности расстояний (pLDDT) AlphaFold 3. Визуализация пространственных моделей вирусных белков выполнена с помощью программы UCSF Chimera.
Синтетические кодон-оптимизированные ДНК копии генов, кодирующих NP SGLV, NP YEZV и NP BJNV, получали путем последовательного отжига частично

комплементарных олигонуклеотидов с последующей амплификацией методом ПЦР и клонированием в плазмиду pJET1.2/blunt. Для получения генетических конструкций, обеспечивающих синтез YEZV NP, SGLV NP и BJNV NP, производили амплификацию ДНК копий генов YEZV NP, SGLV NP и BJNV NP из pJET1.2/YEZV-NP, pJET1.2/SGLV-NP и pJET1.2/BJNV-NP с последующей обработкой ПЦР-продуктов эндонуклеазой рестрикции MalI. Рекомбинантные плазмиды, кодирующие YEZV NP, SGLV NP и BJNV NP, трансформировались в штамм KRX *Escherichia coli* и сверхэкспрессировались совместно со слитым белком убиквитин-подобного модификатора (SUMO) и сайтом протеолиза пикорнаином 3C риновируса A28 [5]. Очистку рекомбинантных белков осуществляли методом металл-хелатной аффинной хроматографии с использованием сорбента HisPur Ni-NTA Resin на хроматографической установке Bio-Lab 100. Анализ фракций рекомбинантных белков проводили методом SDS-PAGE. Концентрацию белка в препарате измеряли на флуориметре MAXLIFE Fluorimeter 2.0. Измерение динамического светорассеяния (DLS) от полученных биомолекул проводилось на анализаторе BeNano 180 Zeta Pro.

Результаты: Исследуемые белки NP SGLV, NP YEZV и NP BJNV не имеют известных аналогов с решенной пространственной структурой в PDB. Уровень идентичности нуклеопротеинов этих вирусов с белками с известными структурами составил около 30 %. С помощью AlphaFold 3 были получены модели структур NP SGLV, NP YEZV и NP BJNV с коэффициентами достоверности pLDDT NP SGLV = 86.45, pLDDT NP YEZV = 86.85, pLDDT NP BJNV = 81.30. Все модели являются высокоструктурированными и имеют два пространственно разделенных домена, в основном представленных альфа-спиралями и соединенных между собой свободными петлями, что указывает на их пригодность для кристаллизации и РСА. Модели белков NP SGLV, NP YEZV и NP BJNV демонстрируют высокий уровень топологического сходства между собой и ортонайровирусом ККГЛ, несмотря на отсутствие гомологов первичной структуры. Пространственная организация белков напоминает ракету, состоящую из домена плеча «стебля» (домен 1) и глобулярного домена «головы» (домен 2) (рис. 1). Распределение электростатического потенциала на поверхности нуклеопротеинов демонстрирует наличие РНК-связывающего сайта в районе центральной щели и в домене 1.

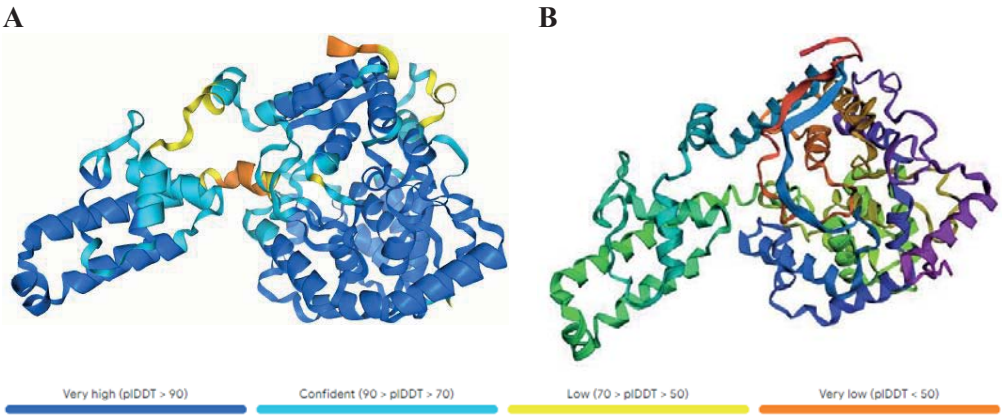


Рис. 1. Модели пространственной структуры мономера нуклеопротеина YEZV. A – структура изображена в цветовом кодирование коэффициента pLDDT; B – структура в ленточном представлении

На втором этапе работы были получены *de novo* ДНК копии генов, кодирующих NP SGLV (1458 нуклеотидов), NP YEZV (1509 нуклеотидов) и NP BJNV (1659 нуклеотидов). Дизайн генетических конструкций производили таким образом, чтобы получить полноразмерные нуклеопротеины и отдельно их домены в растворимой форме. Из нуклеопротеинов были исключены неструктурированные со стороны C и N концов области. Далее после наработки биомассы и двух последовательных хроматографических очисток были получены рекомбинантные NP YEZV молекулярной массой $56.2 \cdot 10^3$ Да, домен 1 NP YEZV – $14.2 \cdot 10^3$ Да, NP BJNV – $61 \cdot 10^3$ Да и NP SGLV – $54.2 \cdot 10^3$ Да без аффинных меток и белков слияния с чистотой более 95 %, пригодные для дальнейших структурных исследований экспериментальными методами, в частности PCA. Методом DLS установлено, что размер биомолекулы рекомбинантного домена 1 NP YEZV варьирует от 2.26 до 3.84 нм, что соответствует ожидаемому. В настоящее время ведутся работы по подбору начальных условий кристаллизации NP SGLV, NP YEZV, домена 1 NP YEZV и NP BJNV.

Выводы: В ходе данного исследования на основе вычислительных моделей были выявлены структурные особенности нуклеопротеинов трех новых ортанайровирусов, а также разработаны подходы для экспериментального исследования структуры рекомбинантных белков малоизученных потенциально патогенных для человека вирусов. Дальнейшее исследование структуры белков новых вирусов даст представление о фундаментальных механизмах инфекций, вызванных данными вирусами, и позволит таргетно подходить к разработке эффективных вакцин и противовирусных препаратов для обеспечения национальной биологической безопасности.

Финансирование: Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (соглашение от 12.10.2021 № 075-15-2021-1355) в рамках реализации отдельных мероприятий Федеральной научно-технической программы развития синхротронных и нейтронных исследований и исследовательской инфраструктуры на 2019–2027 гг.

Structural features of nucleoproteins of the novel orthonairoviruses: Yezo virus, Songling virus and Beiji nairovirus

Gladysheva A.*, Tuyrin V., Tsishevskaya A., Imatdinov I., Agafonov A.

State Research Center of Virology and Biotechnology "Vector" Rospotrebnadzor, Koltsovo, Russia

** gladysheva_av@vector.nsc.ru*

Key words: RNA viruses; *Orthonairovirus*; viral proteins; nucleoprotein

Motivation and Aim: Viruses cause serious illness throughout the world, especially in developing countries. As populations grow in size, density, and mobility, the risk of outbreaks of known and unknown viral infections increases. In recent years, using NGS technology, many new tick-borne viruses pathogenic to humans have been identified, including orthonairoviruses: Songling virus (SGLV) [1], Yezo virus (YEZV) [2, 3] and Beiji nairovirus (BJNV) [4]. The limited data available on these viruses indicate that they are highly likely to cause fever in humans, representing an emerging public health

threat. At the same time, information about the structural features of the virion and proteins of new orthonairoviruses, which is necessary for the rational creation of effective vaccines and therapeutic antiviral drugs, is currently completely absent. Therefore, the goal of this work was to study the structural features of the nucleoproteins of SGLV, YEZV and BJNV, which play an important role in the life cycle of orthonairoviruses and contain immunodominant regions.

Methods and Algorithms: Full-length segment 1 sequences of SGLV, YEZV, and BJNV were obtained from the GenBank database. A search for closely related nucleoproteins whose spatial structures have already been solved experimentally was performed using NCBI BLAST (National Center for Biotechnology Information Basic Local Alignment Search Tool) against the PDB database using the blastp (protein−protein BLAST) algorithm. Multiple alignment of amino acid sequences was performed in the Unipro UGENE program using the MUSCLE iterative alignment method. This program was also used to search and map amino acid substitutions and analyze their location. Functional annotation of viral proteins was performed using Protomenal programs and HMMER wed-service. The search for motifs was carried out using the Pfam and NCBI CDD (Conserved Domains Database) databases. Modeling of the NP SGLV, NP YEZV, and NP BJNV structures was carried out on the AlphaFold server. The selection of spatial models for further theoretical analysis was based on the confidence coefficient for each amino acid considering the predicted per-residue model confidence score (pLDDT) AlphaFold 3. Visualization of spatial structures of viral proteins was performed using the UCSF Chimera program.

Synthetic codon-optimized DNA copies of the genes encoding SGLV NP, YEZV NP, and BJNV NP were obtained by sequential annealing of partially complementary oligonucleotides, followed by PCR and cloning into the pJET1.2/blunt plasmid. DNA copies of the YEZV NP, SGLV NP and BJNV NP genes were amplified from pJET1.2/YEZV-NP, pJET1.2/SGLV-NP, and pJET1.2/ BJNV-NP, followed by treatment of PCR products with restriction endonuclease MalI in order to obtain genetic constructs providing the synthesis of YEZV NP, SGLV NP, and BJNV NP. Recombinant plasmids encoding YEZV NP, SGLV NP, and BJNV NP were transformed into *Escherichia coli* strain KRX and overexpressed together with a small ubiquitin-like modifier (SUMO) fusion protein and the picornaine 3C proteolysis site of rhinovirus A28 [5]. Purification of recombinant proteins was carried out by metal chelate affinity chromatography using HisPur Ni-NTA Resin sorbent on a Bio-Lab 100 chromatography unit. Analysis of recombinant protein fractions was carried out by SDS-PAGE. The protein concentration in the preparation was measured using a MAXLIFE Fluorimeter 2.0. Dynamic light scattering (DLS) measurements from the obtained biomolecules were carried out on a BeNano 180 Zeta Pro analyzer.

Results: The SGLV NP, YEZV NP, and BJNV NP proteins have no known analogues with a solved spatial structure in the PDB. The identity level of the nucleoproteins of these viruses with proteins with known structures in the PDB was about 30 %. Models of the SGLV NP, YEZV NP, and BJNV NP structures were obtained with confidence coefficients: pLDDT NP SGLV = 86.45, pLDDT NP YEZV = 86.85, pLDDT NP BJNV = 81.30 using AlphaFold 3. All models are highly structured and have two spatially separated domains, mainly represented by α-helices and interconnected by free loops, which indicates their suitability for crystallization and X-ray diffraction analysis. The SGLV NP, YEZV NP, and BJNV NP protein models demonstrate a high level of topological similarity between themselves and Crimean–Congo hemorrhagic fever virus,

despite poor primary sequence similarity. The spatial organization of proteins resembles a racket-shaped overall structure, consisting of a “stalk” arm domain (domain 1) and a globular “head” domain (domain 2) (Fig. 1). The distribution of electrostatic potential on the surface of SGLV NP, YEZV NP, and BJNV NP demonstrates the presence of an RNA-binding site in the central cleft region and in the domain 1.

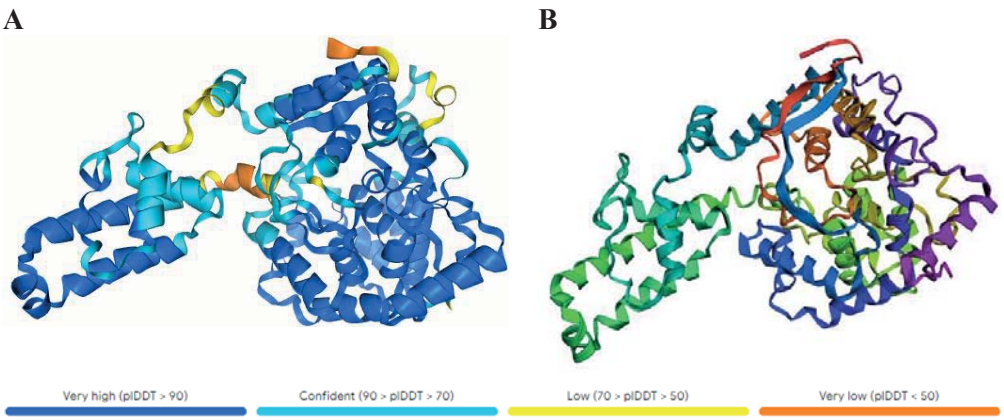


Fig. 1. Structure of monomer YEZV NP.
A – overall structure in pLDDT color representation; B – overall structure in cartoon representation

At the second stage of the work, *de novo* DNA copies of the genes encoding SGLV NP (1458 nucleotides), YEZV NP (1509 nucleotides), and BJNV NP (1659 nucleotides) were obtained. The design of genetic constructions was carried out in such a way as to obtain full-length nucleoproteins and separately their domains in soluble form. Unstructured regions at the C and N ends of SGLV NP, YEZV NP, and BJNV NP were excluded from the recombinant nucleoproteins. Then, recombinant YEZV NP with a molecular weight of $56.2 \cdot 10^3$ Da, domain 1 YEZV NP – $14.2 \cdot 10^3$ Da, BJNV NP – $61 \cdot 10^3$ Da and SGLV NP – $54.2 \cdot 10^3$ Da were obtained after production of biomass and two sequential chromatographic purifications. Recombinant SGLV NP, YEZV NP, and BJNV NP had not contain affinity tags and fusion proteins, had a purity of more than 95 %, and were suitable for further structural studies by experimental methods, in particular X-ray diffraction. Using the DLS method, it was found that the size of the biomolecule of the recombinant domain 1YEZV NP varies from 2.26 to 3.84 nm, which corresponds to the expected value. Currently, research is underway to select the initial conditions of SGLV NP, YEZV NP, domain 1 YEZV NP, and BJNV NP crystallization.
Conclusion: In the course of this study, based on computational models, the structural features of the nucleoproteins of three new orthonairoviruses were identified, and approaches were developed for the experimental study of the structure of recombinant proteins of poorly studied viruses that are potentially pathogenic for humans. Further study of the protein structure of new viruses will provide insight into the fundamental mechanisms of infections caused by these viruses and will allow for a targeted development of effective vaccines and antiviral drugs to ensure national biological safety.

Funding: The study was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2021-1355 dated October 12, 2021) as a part of the implementation of certain activities of the Federal Scientific and Technical Program for the Development of Synchrotron and Neutron Research and Research Infrastructure for 2019–2027.

Список литературы/References

1. Ma J., Lv X.L. et al. Identification of a new orthonairovirus associated with human febrile illness in China. *Nat Med*. 2021;27(3):434-439. doi 10.1038/s41591-020-01228-y
2. Kodama F., Yamaguchi H. et al. A novel nairovirus associated with acute febrile illness in Hokkaido, Japan. *Nat Commun*. 2021;12(1):5539. doi 10.1038/s41467-021-25857-0
3. Lv X., Liu Z. et al. Yezo Virus Infection in Tick-Bitten Patient and Ticks, Northeastern China. *Emerg Infect Dis*. 2023;29(4):797-800. doi 10.3201/eid2904.220885
4. Wang Y.C., Wei Z. et al. A new nairo-like virus associated with human febrile illness in China. *Emerg Microbes Infect*. 2021;10(1):1200-1208. doi 10.1080/22221751.2021.1936197
5. Tishin A.E., Gladysheva A.V., Pyatavina L.A. et al. Preparation and Crystallization of Picornain 3C of Rhinovirus A28. *Crystallogr Rep*. 2023;68:924-930. doi 10.1134/S1063774523601119

doi 10.18699/bgrs2024-3.1-07

Поиск потенциальных сайтов связывания полифенольных соединений в структуре рибонуклеотидредуктазы вируса герпеса человека 1 типа

Дубинкина Е.С.[1, 3]*, Галицкая А.А.[1], Анашкина А.А.[2, 3]

[1] *Саратовский национальный исследовательский государственный университет им. Н.Г. Чернышевского, Саратов, Россия*

[2] *Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

[3] *Цифровая кафедра Института цифровой медицины Сеченовского университета, Москва, Россия*

* *kometa.ed@gmail.com*

Ключевые слова: герпес-вирусная инфекция; ВПГ1; полифенольные соединения; рибонуклеотидредуктаза; молекулярный докинг; сайты связывания

Мотивация и цель: Герпетическая инфекция, вызываемая ДНК-содержащими вирусами семейства Herpesviridae, несмотря на бесчисленное количество попыток установления ее молекулярных патогенетических механизмов, остается объектом активных научных исследований и по сей день. Попав однажды в организм, инфекция сохраняется в нем навсегда и может латентно персистировать на протяжении всей жизни в нейронах инфицированных людей [1]. Наиболее распространенным из девяти видов [2] считается вирус простого герпеса 1 типа (ВПГ1), который находится в неактивном состоянии, а при активации становится причиной орального и генитального герпеса, конъюнктивита, вторичного бесплодия, неонатальной гибели ребенка [3]. Прорывом в лечении этих инфекций стало создание в 1980-х гг. первого противовирусного препарата «Ацикловир», синтетического ациклического аналога гуанозина, который является одним из самых частых концевых и внутренних нуклеозидов ДНК герпес-вирусов, что обусловило очень высокую терапевтическую активность этого препарата. Однако сегодня рост числа резистентных к ацикловиру штаммов ВПГ1 указывает на необходимость разработки новых противовирусных средств для борьбы с герпес-вирусами. 1976-й год стал датой, когда было проведено одно из первых исследований, продемонстрировавших противовирусную активность полифенольных соединений [4]. На сегодняшний день уже существуют доказательства выраженного действия полифенолов против различных вирусов: гриппа, Зика, гепатита B и C, лихорадки западного Нила, кори [5], а также против вируса герпеса [1, 6]. Поэтому направление поиска новых соединений для терапии и профилактики ВПГ1, которые бы эффективно влияли на факторы патогенеза ВПГ1, обладая при этом профилем безопасности, и, более того, объяснение механизмов действия этих соединений являются актуальной задачей данной области науки, требующей усилий химиков, биологов, фармацевтов и медиков. Целью настоящей работы был поиск потенциальных сайтов связывания полифенольных соединений в структуре рибонуклеотидредуктазы ВПГ1.

Методы и алгоритмы: Выбор рибонуклеотидредуктазы в качестве мишени был обусловлен тем, что данный белок ответственен за реактивацию ВПГ1 из латентного состояния в покоящихся нейронах посредством синтеза дезоксипиримидинов

из рибопиримидинов [2]. Полноатомная 3D модель вирусной рибонуклеотидредуктазы была получена путем построения гомологичных моделей на основе структуры человеческого белка p53R2 (3HF1) и последовательности вирусной рибонуклеотидредуктазы (P06474) в программе MOE. В качестве исследуемых полифенольных соединений были выбраны следующие вещества: цис-ресвератрол, транс-ресвератрол, кверцетин, пицеатанол, куркумин, подофилотоксин, ладанеин, хонокиол, пеларгонидин, розинидин, трицетинидин [5]. Для выбранных полифенольных соединений к полученной модели вирусной рибонуклеотидредуктазы была проведена процедура молекулярного докинга.

Результаты: Полифенольные соединения показали наличие 14 возможных сайтов связывания в структуре рибонуклеотидредуктазы ВПГ1 (рис. 1).

Наилучшую афинность связывания (–6.5 ккал/моль) продемонстрировал куркумин, потенциально взаимодействующий с аминокислотными остатками Glu187, Gly188, Phe190, Phe191, Ala192, Phe195, Ala196, Ser218, Glu221, Ala222, Thr225, Thr226, Tyr284, Ser288, Arg291, Leu292, Leu295, Phe311 (табл. 1 и 2).

Таблица 1. Потенциальные сайты связывания в структуре рибонуклеотидредуктазы ВПГ1

№	Size	PLB	Hyd	Side	Residue
2	16	–0.36	8	10	Glu77, Gly78, Ser81, Phe82, Phe85, Gln265, Ala266, Pro267, Leu273
4	29	0.09	22	34	Glu77, Ser81, Phe82, Arg84, Phe85, Arg146, Tyr149, Val150, Thr153, Gln265, Ala266, Pro267, Leu273
5	20	–0.94	6	8	Leu8, Pro10, Gly17, His18, Ser19, Leu236, Gly237, Gly238
8	62	2.20	17	37	1 цепь: Pro34, Glu35, Arg36, Tyr37, Gly102, Ser105, Gly106, Leu107, Phe108, Glu109, Lys111, Leu114 2 цепь: Ser91, Ala92, Ala93, Asp95, Leu96, Thr98, Glu99, Gly102, Glu121, Asn132
9	25	–0.5	6	14	1 цепь: Val97, Glu99, Gly102, Gly103, Glu121 2 цепь: Gly102, Ser105, Gly106, Leu114
10	24	–0.19	7	13	Pro306, Pro307, Asp308, Ala309, Leu313, Ser314, Leu315, Met316, Ser317, Lys320
11	76	2.85	26	41	Glu187, Gly188, Phe190, Phe191, Ala192, Phe195, Ala196, Ser218, Glu221, Ala222, Thr225, Thr226, Tyr284, Ser288, Arg291, Leu292, Leu295, Phe311
12	11	–0.74	3	11	Ala199, Arg202, Thr203, Asp215, Ser310

Примечание. № – номер сайта связывания; Size – количество α-субъединиц, составляющих сайт; PLB – показатель склонности к связыванию лиганда [7], Hyd – количество гидрофобных контактных атомов в рецепторе, Side – количество контактных атомов боковой цепи в рецепторе, Residue – аминокислотные остатки.

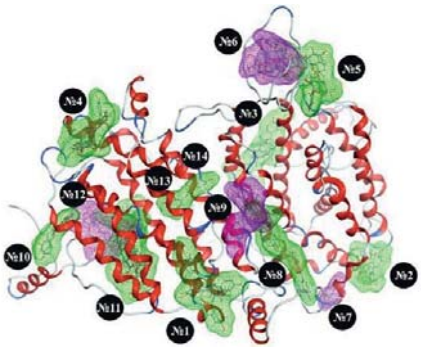


Рис. 1. Потенциальные сайты связывания полифенольных соединений в структуре рибонуклеотидредуктазы ВПГ1

Следующий по афинности результат (–5.9 ккал/моль) показал подофилотоксин, потенциально взаимодействующий с аминокислотами Glu77, Gly78, Ser81, Phe82, Phe85, Gln265, Ala266, Pro267, Leu273. Третий по афинности результат показал хонокиол (–5.7 ккал/моль), взаимодействуя в том же сайте, что и куркумин (см. табл. 1 и 2).

Таблица 2. Взаимодействие полифенольных соединений с потенциальными сайтами связывания и значение энергии связи

Название соединения	Номер сайта связывания	Энергия связи, ккал/моль
Цис-ресвератрол	12	–4.4
Транс-ресвератрол	2	–5.1
Кверцетин	4	–5.2
Пицеатанол	9	–5.1
Куркумин	11	–6.6
Подофилотоксин	2	–5.9
Хонокиол	11	–5.8
Ладанеин	11	–5.4
Пеларгонидин	8	–5.3
Розинидин	5	–5.5
Трицетинидин	10	–5.2

Выводы: Предложены потенциальные сайты связывания полифенольных соединений куркумина, подофилотоксина и хонокиола с белком рибонуклеотидредуктазой ВПГ1 для последующей экспериментальной проверки.
Финансирование: Исследование поддержано грантом Российского научного фонда № 24-24-00563 на тему «Разработка цифровых образовательных программ в биомедицине».

Search for potential binding sites for polyphenolic compounds in the structure of human herpes virus type 1 ribonucleotide reductase

Dubinkina E.S.[1, 3]*, Galitskaya A.A.[1], Anashkina A.A.[2, 3]

[1] *Saratov National Research State University named after N.G. Chernyshevsky, Saratov, Russia*

[2] *Institute of Molecular Biology named after. V.A. Engelhardt, RAS, Moscow, Russia*

[3] *Digital Department, Institute of Digital Medicine, Sechenov University, Moscow, Russia*

* *kometa.ed@gmail.com*

Key words: herpes viral infection; HSV1; polyphenolic compounds; ribonucleotide reductase; molecular docking; binding sites

Motivation and Aim: Herpes infection, caused by DNA-containing viruses of the Herpesviridae family, despite countless attempts to establish its molecular pathogenetic mechanisms, remains the object of active scientific research to this day. Once the infection enters the body, it remains there forever and can persist latently throughout life in the neurons of infected people [1]. The most common of the nine types of this virus [2] is considered to be herpes simplex virus type 1 (HSV1), which is in an inactive state, and when activated becomes the cause of oral and genital herpes, conjunctivitis, secondary infertility, neonatal death of a child [3]. A breakthrough in the treatment of these infections was the creation in the 1980s of the first antiviral drug "Acyclovir", a synthetic acyclic analogue of guanosine, which is one of the most common terminal and internal nucleosides of the DNA of herpes viruses, which led to the very high therapeutic activity of this drug. However, today, the increase in the number of acyclovir-resistant HSV1 strains indicates the need for the development of new antiviral agents to combat herpes viruses. 1976 was the date when one of the first studies was conducted to demonstrate the antiviral activity of polyphenolic compounds [4]. Today, there is already evidence of a pronounced effect of polyphenols against various viruses: influenza, Zika, hepatitis B and C, West Nile fever, measles [5], as well as against the herpes virus [1, 6]. Therefore, the search for new compounds for the treatment and prevention of HSV1, which would effectively influence the factors of HSV1 pathogenesis, while having a safety profile, and, moreover, explaining the mechanisms of action of these compounds is a very urgent task in this area of science and requires the efforts of chemists, biologists, pharmacists and doctors. The purpose of this work was to search for potential binding sites for polyphenolic compounds in the structure of HSV1 ribonucleotide reductase.
Methods and Algorithms: The choice of ribonucleotide reductase as a target was due to the fact that this protein is responsible for the reactivation of HSV1 from a latent state in resting neurons through the synthesis of deoxypyrimidines from ribopyrimidines [2]. An all-atomic 3D model of viral ribonucleotide reductase was obtained by constructing homology models based on the structure of the human p53R2 (3HF1) protein and the sequence of viral ribonucleotide reductase (P06474) in the MOE program. Cis-resveratrol, trans-resveratrol, quercetin, piceataanol, curcumin, podophyllotoxin, honokiol, ladanein, pelargonidin, rosinidin, tricetinidin [5] were selected from all the polyphenolic compounds as study compounds. For selected polyphenolic compounds, a molecular docking procedure to the resulting model of viral ribonucleotide reductase was carried out.

Results: Polyphenolic compounds showed the presence of 14 possible binding sites in the structure of HSV1 ribonucleotide reductase (Fig. 1).

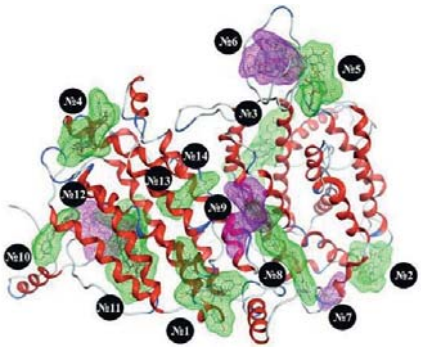


Fig. 1. Potential binding sites for polyphenolic compounds in the structure of HSV1 ribonucleotide reductase

The best binding affinity (–6.5 kcal/mol) was demonstrated by curcumin, which potentially interacts with amino acid residues Glu187, Gly188, Phe190, Phe191, Ala192, Phe195, Ala196, Ser218, Glu221, Ala222, Thr225, Thr226, Tyr284, Ser288, Arg291, Leu292, Leu295, Phe311 (Table 1, Table 2).

Table 1. Potential binding sites in the structure of HSV1 ribonucleotide reductase

No.	Size	PLB	Hyd	Side	Residue
2	16	–0.36	8	10	Glu77, Gly78, Ser81, Phe82, Phe85, Gln265, Ala266, Pro267, Leu273
4	29	0.09	22	34	Glu77, Ser81, Phe82, Arg84, Phe85, Arg146, Tyr149, Val150, Thr153, Gln265, Ala266, Pro267, Leu273
5	20	–0.94	6	8	Leu8, Pro10, Gly17, His18, Ser19, Leu236, Gly237, Gly238
8	62	2.20	17	37	1 chain: Pro34, Glu35, Arg36, Tyr37, Gly102, Ser105, Gly106, Leu107, Phe108, Glu109, Lys111, Leu114 2 chain: Ser91, Ala92, Ala93, Asp95, Leu96, Thr98, Glu99, Gly102, Glu121, Asn132
9	25	–0.5	6	14	1 chain: Val97, Glu99, Gly102, Gly103, Glu121 2 chain: Gly102, Ser105, Gly106, Leu114
10	24	–0.19	7	13	Pro306, Pro307, Asp308, Ala309, Leu313, Ser314, Leu315, Met316, Ser317, Lys320
11	76	2.85	26	41	Glu187, Gly188, Phe190, Phe191, Ala192, Phe195, Ala196, Ser218, Glu221, Ala222, Thr225, Thr226, Tyr284, Ser288, Arg291, Leu292, Leu295, Phe311
12	11	–0.74	3	11	Ala199, Arg202, Thr203, Asp215, Ser310

Note. No. – binding site number; Size – number of α-subunits making up the site; PLB is an indicator of the propensity to bind a ligand [7], Hyd is the number of hydrophobic contact atoms in the receptor, Side is the number of contact atoms of the side chain in the receptor, Residue is amino acid residues.

The next highest affinity result (–5.9 kcal/mol) was shown by podophyllotoxin, potentially interacting with the amino acids Glu77, Gly78, Ser81, Phe82, Phe85, Gln265,

Ala266, Pro267, Leu273. The third highest affinity result was shown by honokiol (–5.7 kcal/mol), interacting at the same site as curcumin (Table 1, Table 2).

Table 2. Interaction of polyphenolic compounds with potential binding sites and binding energy values

Compound's name	Binding site number	Binding energy, kcal/mol
Cis-resverantrol	12	–4.4
Trans-resverantrol	2	–5.1
Quercetin	4	–5.2
Piceataanol	9	–5.1
Curcumin	11	–6.6
Podophyllotoxin	2	–5.9
Honokiol	11	–5.8
Ladanein	11	–5.4
Pelargonidin	8	–5.3
Rosinidin	5	–5.5
Tricetinidin	10	–5.2

Conclusion: Potential binding sites for the polyphenolic compounds curcumin, podophyllotoxin and honokiol with the HSV1 ribonucleotide reductase protein were proposed for subsequent experimental testing.

Funding: The research was supported by a grant from the Russian Science Foundation No. 24-24-00563 on the topic “Development of digital educational programs in biomedicine”.

Список литературы/References

1. Annunziata G., Maisto M., Schisano C., Ciampaglia R., Narciso V., Tenore G.C., Novellino E. Resveratrol as a Novel Anti-Herpes Simplex Virus Nutraceutical Agent: An Overview. *Viruses*. 2018;10(9):473. doi 10.3390/v10090473
2. Куханова М.К., Коровина А.Н., Кочетков С.Н. Вирус простого герпеса человека: жизненный цикл и поиск ингибиторов. *Успехи биологической химии*. 2014;54:457-494
3. Викулов Г.Х. Иммунологические аспекты герпес-вирусных инфекций. *Клиническая дерматология и венерология*. 2015;14(5):104-116
4. Лопатина А.В., Кукушкина А.Д., Мельников М.В., Роговский В.С. Перспективы применения полифенолов при рассеянном склерозе. *Журнал неврологии и психиатрии*. 2022;122(7/2):36-43
5. Яшин А.Я., Веденин А.Н., Яшин Я.И., Василевич Н.И. Антивирусные полифенолы – антиоксиданты: структура, пищевые источники и механизм действия. *Лаборатория и производство*. 2020;5(14):76-86
6. El-Toumy S.A., Salib J.Y., El-Kashak W.A. et al. Antiviral effect of polyphenol rich plant extracts on herpes simplex virus type 1. *Food Sci Hum Wellness*. 2018;7:91-101
7. Soga S., Shirai H., Kobori M., Hirayama N. Use of Amino Acid Composition to Predict Ligand-Binding Sites. *J Chem Inf Model*. 2007;47(2):400-406

doi 10.18699/bgrs2024-3.1-08

Структура кристаллов комплекса ДНК с гистоноподобным белком Dps

Коваленко В.В.[1]*, Лойко Н.Г.[2], Терешкин Э.В.[1], Терешкина К.Б.[1], Крупянский Ю.Ф.[1]

[1] *Федеральный исследовательский центр химической физики им. Н.Н. Семёнова РАН, Москва, Россия*

[2] *Федеральный исследовательский центр «Фундаментальные основы биотехнологии» РАН, Москва, Россия*

* *vladislavkovalenko785@gmail.com*

Ключевые слова: просвечивающая электронная микроскопия; кристаллизация ДНК; молекулярное моделирование

Мотивация и цель: Бактериальные клетки в ходе эволюции выработали универсальные наследственные стратегии адаптации, основанные на структурных, биохимических и генетических перестройках, позволяющие сохранить часть популяции и выжить в любых неблагоприятных условиях. Бактериальные стратегии адаптации в первую очередь направлены на защиту ДНК с помощью нуклеоид- ассоциированных белков [1]. Так, при приближении к стационарной фазе роста в бактериях запускается молекулярный механизм выработки (до 200 000 единиц) ферретиноподобного белка Dps (DNA-binding protein of starved cells). Белок Dps участвует в конденсации ДНК, уплотняя ее в более высокоупорядоченную «биокристаллическую» или «фибро-нанокристаллическую» структуру, максимально защищающую от разрушительных воздействий. Несмотря на высокий интерес со стороны научного сообщества к данной теме, до сих пор не получено точной молекулярной структуры комплекса ДНК-Dps. Однако решение этого вопроса важно, так как прольет свет на механизмы выживания бактерий в стрессовых условиях и позволит подобрать новые способы стерилизации в пищевой промышленности и медицине.

Методы и алгоритмы: Для определения структуры макромолекулярных комплексов и молекулярных механизмов взаимодействия ДНК с Dps использованы методы просвечивающей электронной микроскопии, дополненные инструментами молекулярного моделирования.

Результаты: Кристаллические *in vitro* комплексы Dps-ДНК были получены смешиванием очищенного белка Dps (3.4 мг/мл) с линейным двуцепочечным фрагментом ДНК (0.25 нг/мкл) pBluescript II SK(+) длиной 2961 пар оснований непосредственно на покрытую углеродом медную сетку ТЭМ (SPI) с последующим немедленным добавлением ЭДТА (0.14 мМ). Таким образом, кристаллы формировались непосредственно на сетке ПЭМ. Снимки ПЭМ содержали множество кристаллов линейного размера от 50 до 500 нм (рис. 1). Томограммы были получены с использованием программного обеспечения Jeol для томографии (Jeol, Япония). Угол наклона гониометра колеблется от –60 до +60 (с постоянным шагом 1 градус). Серия изображений была выровнена с помощью

программного обеспечения Digital Micrograph (Gatan, США), а затем восстановлена с помощью алгоритма обратной проекции в IMOD4.11. Томография содержала периодические области электронной плотности, с пространственной группой P1 и постоянными элементарной ячейки a = 84.1 Å, b = 77.0 Å, c = 45.0 Å, α = 90.00°, β = 90.00°, γ = 113.30°. Белок Dps в додекамерной четвертичной форме (сферическая форма приблизительно 90 Å в диаметре) [2] не может разместиться в данной элементарной ячейке. Сама электронная плотность внутри элементарной ячейки имела форму пирамиды с полостью внутри в виде усеченного конуса, что по виду совпадала с частью додекамерной формы белка Dps в области соприкосновения трех мономеров Dps, образующих так называемую Dps пору. Для подтверждения существования белка в тримерной четвертичной структуре, методами молекулярного моделирования в программном пакете Gromacs 5.1 [3] в полноатомном силовом поле AMBER был проведен расчет трех мономеров белка Dps в водном растворе в течение 1 микросекунды. По результатам расчетов была вычислена потенциальная энергия взаимодействия мономеров (–185.98 кДж/моль). Далее отрелаксированный тример и молекулы ДНК были размещены в электронную плотность кристалла с помощью программы UCSF Chimera [4] (рис. 2).

Выводы: В нашем эксперименте была впервые обнаружена структура белка Dps в четвертичной форме, отличной от додекамерной, внутри кристаллических образований очень малого размера. Подобные структуры также были обнаружены нами в покоящихся клеточных образцах *E. coli* [5], где размер периодической решетки внутриклеточных кристаллов был меньше, чем диаметр додекамера Dps. По-видимому, белок Dps в клетках, защищенных от внешних стрессовых воздействий также находится в тримерной четвертичной структуре.

Финансирование: Исследование проведено при финансовой поддержке гранта РНФ № 23-24-00250.

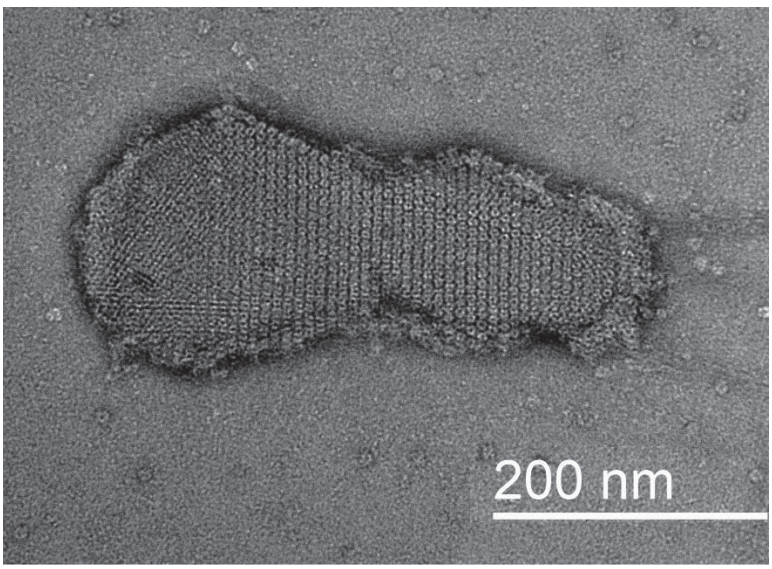


Рис. 1. ПЭМ снимок кристаллического комплекса ДНК-Dps

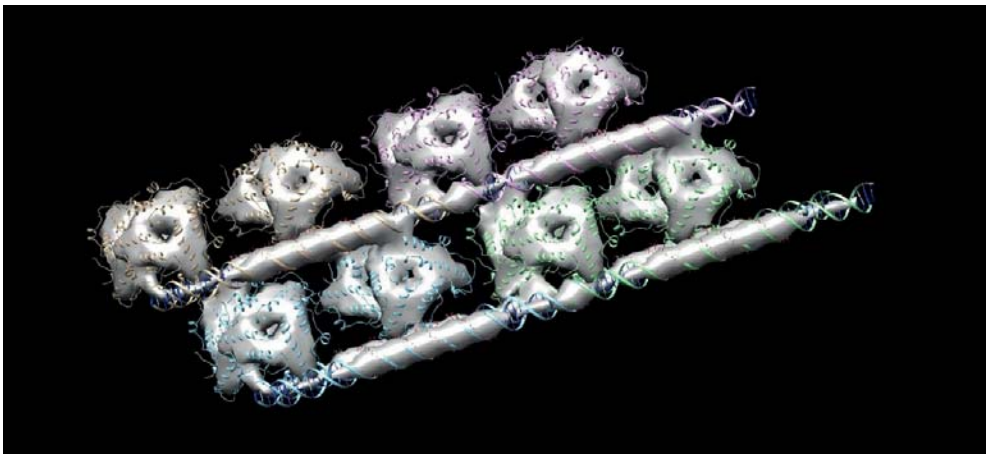


Рис. 2. Финальное расположение тримеров белка Dps и молекул ДНК в карту электронной плотности, полученную из томографии

Crystal structure of DNA with histone-like protein Dps complex

Kovalenko V.V.[1]*, Loiko N.G.[2], Tereshkin E.V.[1], Tereshkina K.B.[1], Krupyanskii Yu.F.[1]

[1] *N.N. Semenov Federal Research Center for Chemical Physics, RAS, Moscow, Russia*

[2] *Federal Research Center "Fundumentals of Biotechnology", RAS, Moscow, Russia*

** vladislavkovalenko785@gmail.com*

Key words: transmission electron microscopy; DNA crystallization; molecular modeling

Motivation and Aim: Evolutionary bacterial cells evolution have developed universal hereditary adaptation strategies based on structural, biochemical, and genetic rearrangements that allow them to conserve a part of the population and survive in any unfavorable conditions. Bacterial adaptation strategies are primarily aimed at protecting DNA with the help of nucleoid-associated proteins [1]. Thus, as bacteria approach the stationary phase of growth, a molecular mechanism of production (up to 200,000 units) of the ferretin-like protein Dps (DNA-binding protein of starved cells) is triggered. The Dps protein participates in the condensation of DNA, compacting it into a more highly ordered "biocrystalline" or "fibro-nanocrystalline" structure that maximizes protection from damaging effects. Despite the high interest of the scientific community in this topic, the exact molecular structure of the DNA-Dps complex has not yet been obtained. However, the solution of this issue is important, as it will shed light on the mechanisms of bacterial survival under stress conditions and allow to select new methods of sterilization in the food industry and medicine.
Methods and Algorithms: Transmission electron microscopy techniques supplemented with molecular modeling tools were used to determine the structure of macromolecular complexes and molecular mechanisms of interaction between DNA and Dps.
Results: Crystalline in vitro Dps-DNA complexes were obtained by mixing purified Dps protein (3.4 mg/mL) with a linear double-stranded DNA fragment (0.25 ng/μL) of pBluescript II SK(+) with a length of 2961 base pairs directly onto a carbon-coated

copper TEM grid (SPI) followed by immediate addition of EDTA (0.14 mM). Thus, crystals were formed non-intermediately on the PEM grid. TEM images contained many crystals with linear sizes ranging from 50 to 500 nm (Fig. 1). Tomograms were obtained using Jeol tomography software (Jeol, Japan). The goniometer tilt angle ranged from –60 to +60 (with a constant step of 1 degree). The image series were aligned using Digital Micrograph software (Gatan, USA) and then reconstructed using the back-projection algorithm in IMOD4.11. The tomography contained periodic regions of electron density, with space group P1 and unit cell constants equal to a = 84.1 Å, b = 77.0 Å, c = 45.0 Å, $\alpha = 90.00°$, $\beta = 90.00°$, $\gamma = 113.30°$. The Dps protein in the dodecameric quaternary form (spherical shape approximately 90 Å in diameter) [2] cannot be fit into this unit cell. The very electron density inside the unit cell had the shape of a pyramid with a cavity inside in the form of a truncated cone, which coincides with the part of the dodecameric form of the Dps protein in the region of contact of three Dps monomers forming the so-called Dps pore. To confirm the existence of the protein in the trimeric quaternary structure, three monomers of the Dps protein in aqueous solution for 1 microsecond were calculated by molecular modeling methods in the Gromacs 5.1 software package [3] in the all-atom AMBER force field. The potential energy of monomer interaction – 185.98 kJ/mol) was obtained based on the results of the calculated trajectories. Further, the trimer structure after energy relaxation process and DNA molecules were placed in the electron density of the crystal using the UCSF Chimera program [4] (Fig. 2).

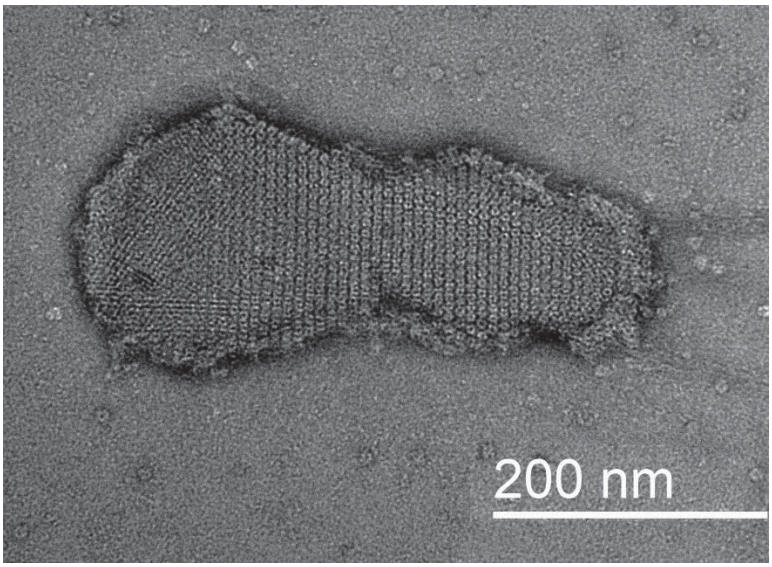


Fig. 1. TEM image of the crystalline complex of DNA-Dps

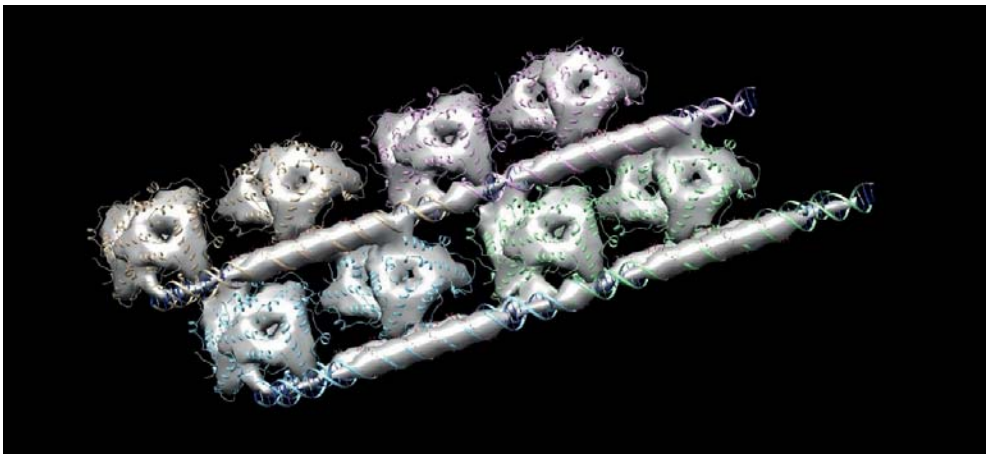


Fig. 2. Final arrangement of Dps protein trimers and DNA molecules in the electron density map obtained from SEM tomography

Conclusion: In our experiment, the structure of Dps protein in the quaternary form different from the dodecamer inside crystalline formations of very small size was detected for the first time. We also found similar structures in resting cell samples of *E. coli* [5], where the size of the periodic lattice of intracellular crystals was smaller than the diameter of the Dps dodecamer. Apparently, the Dps protein in cells protected from external stress effects is also in a trimeric quaternary structure.

Funding: The study was financially supported by RSF grant #23-24-00250.

Список литературы/References

1. Minsky A., Shimoni E., Frenkiel-Krispin D. Stress, order and survival. *Nat Rev Mol Cell Biol*. 2002;3(1):50-60. doi 10.1038/nrm700
2. Grant R.A., Filman D.J., Finkel S.E., Kolter R., Hogle J.M. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol.* 1998;5(4):294-303. doi 10.1038/nsb0498-294
3. Berendsen H.J.C., van der Spoel D., van Drunen R. GROMACS: A message passing parallel molecular dynamics implementation. *Comput Phys Commun.* 1995;91(1):43-56. doi 10.1016/0010-4655(95)00042-E
4. Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E., UCSF Chimera – A visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25(4):1605-1612. doi 10.1002/jcc.20084
5. Loiko N., Danilova Y., Moiseenko A., Kovalenko V., Tereshkina K., Tutukina M., El-Registan G., Sokolova O., Krupyanskii Y. Morphological peculiarities of the DNA-protein complexes in starved *Escherichia coli* cells. *PloS One*. 2020;15(10):e0231562. doi 10.1371/journal.pone.0231562

doi 10.18699/bgrs2024-3.1-09

Методы анализа данных малоуглового рентгеновского рассеяния многокомпонентными биомолекулярными системами

Конарев П.В.
НИЦ «Курчатовский институт», Москва, Россия
konarev@crys.ras.ru

Ключевые слова: малоугловое рентгеновское рассеяние; многокомпонентные биомолекулярные системы; пространственная форма частиц; гель-хроматография; эволюционный факторный анализ

Мотивация и цель: В настоящее время большую роль при рациональном конструировании новых лекарственных средств и развитии методов терапии заболеваний играют структурные методы исследования вещества, среди которых особо следует выделить малоугловое рентгеновское рассеяние (МУРР). Метод МУРР позволяет изучать некристаллические образцы различной природы с размерами частиц от 1 до 200 нм и более. Этот метод является неинвазивным, не требует сложной пробоподготовки и дает ценную информацию о структуре и динамике исследуемых объектов в условиях, близких к нативным. До недавнего времени анализ полидисперсных многокомпонентных биомолекулярных систем по данным МУРР был существенным образом ограничен и позволял оценивать лишь общие структурные характеристики, усредненные по ансамблю частиц. Недавние достижения в разработке алгоритмов позволяют гораздо более детально исследовать полидисперсные системы, в частности белковые или липидные многокомпонентные смеси.
Методы и алгоритмы: При наличии известных кривых малоуглового рассеяния от каждого макромолекулярного компонента можно определять их объемные доли в общей смеси [1, 2]. Если структуру компонентов можно аппроксимировать простыми геометрическими телами, то, помимо определения объемных долей компонент, появляется возможность оценивать функции распределения частиц по размерам и характер их межчастичного взаимодействия в случае концентрированных систем [1, 2]. Для липидных многокомпонентных систем, помимо определения функции распределения везикул по размерам, проводится восстановление электронной плотности липидного бислоя внутри везикулы и оценка параметров мультислойной ламеллярной организации везикул [3, 4].
Хотя прямой анализ формы частиц для многокомпонентных смесей сильно затруднен, в ряде случаев это становится возможным благодаря новому подходу прямого определения формы промежуточных компонентов в эволюционирующих системах [5]. Системы, для которых известны начальное и конечное состояния, но которые также содержат неизвестное промежуточное состояние, могут быть непосредственно проанализированы с помощью этого подхода.
Измерения МУРР в сочетании с использованием гель-хроматографической ячейки, когда раствор образца, выходящий из хроматографической колонки, сразу направляется в камеру образца, позволяют разделить во времени вклады

отдельных (чистых и монодисперсных) компонентов, присутствующих в системе, но не всегда удается достичь полного разделения компонентов. Для частично перекрывающихся пиков выхода компонентов смеси предложен новый алгоритм восстановления профилей рассеяния и концентрации компонентов на основе эволюционного факторного анализа (ЭФА) [6]. Установлены границы применимости подхода в зависимости от уровня шума во входных данных МУРР, степени перекрытия концентрационных профилей компонентов и их асимметрии, что важно для его практических применений.

Результаты: С помощью количественного анализа данных малоуглового рассеяния с использованием параметрических методов линейной и нелинейной минимизации удалось: (i) выявить случайно-столкновительный механизм образования комплекса адренодоксина (Adx) с митохондриальным цитохромом с (Cc) [7]; (ii) показать способность фермента люмазинсинтазы самособираться в капсулы размером от 16 до 32 нм, причем их количественный состав зависит как от наличия мутаций белка, так и от изменений физико-химических условий среды [8]; (iii) проанализировать эволюцию структурных параметров везикул DMPC/DPPC в процессе экструзии, позволяющую получать однослойные везикулы с радиусом (30±1) нм [4]; (iv) для смесей, состоящих из вирусных липидов и матричного белка M1 вируса гриппа A, оценить влияние гемагглютинина на формирование липопротеиновых нанодоменов в липидном бислое [9].

Алгоритм восстановления формы частиц для многокомпонентных смесей был успешно применен для изучения процесса образования фибрилл белка инсулина и позволил определить форму промежуточного компонента в процессе их роста, который представляет собой вытянутую структуру размером 20 нм, состоящую из пяти связанных доменов, по размерам близких к мономерам инсулина, что подтверждает гипотезу об олигомерном ядре, являющемся первичной удлиняющей единицей инсулиновых амилоидных фибрилл [5]. С помощью предложенного метода для белка фактора роста нервов была (случай двухкомпонентной смеси) восстановлена трехмерная форма димера по концентрационной серии данных МУРР [5].

С помощью ЭФА алгоритма определено, что для смеси белка пируват-альдолазы ее олигомерными компонентами в растворе являются гексамеры и октамеры [6]. Восстановленные профили рассеяния компонентов хорошо соотносятся с известными кристаллографическими моделями белков.

Выводы: Разработанные методы существенно расширяют возможности анализа многокомпонентных наноразмерных систем методом МУРР и могут быть применены для широкого круга биологических, органических и композитных систем.

Финансирование: Работа проведена в рамках выполнения государственного задания НИЦ «Курчатовский институт».

Methods for the analysis of small-angle X-ray scattering data from multicomponent biomolecular systems

Konarev P.V.

NRC "Kurchatov Institute", Moscow, Russia

konarev@crys.ras.ru

Key words: small-angle X-ray scattering; multicomponent biomolecular systems; 3D shape of particles; gel chromatography; evolutionary factor analysis

Motivation and Aim: Nowadays, structural methods play an important role in rational design of new drugs and development of therapies for diseases, among which small-angle X-ray scattering (SAXS) should be emphasized. SAXS allows studying non-crystalline samples of different nature with particle sizes from 1 to 200 nm and more. This method is non-invasive, does not require special sample preparation and provides valuable information on the structure and dynamics of the investigated objects under nearly physiological conditions. Until recently, the analysis of SAXS data from polydisperse multicomponent biomolecular systems was significantly limited and allowed only the evaluation of general structural characteristics averaged over an ensemble of particles. Recent advances in algorithm development allow one to extract much more detailed structural information for polydisperse systems, in particular protein or lipid multicomponent mixtures.

Methods and Algorithms: In the presence of known small-angle scattering curves from each macromolecular component, it is possible to determine their volume fractions in a mixture [1, 2]. If the structure of the components can be approximated by simple geometric bodies, in addition to determining the volume fractions of the components, it is possible to estimate the particle size distribution functions and the interparticle interactions in the case of concentrated systems [1, 2]. For lipid multicomponent systems, in addition to determining the size distribution function of vesicles, the electron density of the lipid bilayer inside the vesicle is reconstructed and the parameters of the multilayer lamellar organization of vesicles are estimated [3, 4].

Although direct particle shape analysis for multicomponent mixtures is normally not applicable, a new approach to directly determine the shape of intermediate components in evolving systems makes this possible in some cases [5]. Systems for which the initial and final states are known, but which also contain an unknown intermediate state, can be directly analyzed using this approach.

SAXS measurements combined with the use of a size exclusion chromatography (SEC), where the sample solution leaving the chromatographic column is immediately sent to the sample chamber, allow for the separation in time of the contributions of the individual (pure and monodisperse) components present in the system, but it is not always possible to achieve complete separation of the components. For partially overlapping peaks of the mixture components, a new algorithm for reconstructing the scattering and concentration profiles of components based on evolutionary factor analysis (EFA) has been proposed [6]. The limitations of the approach are evaluated depending on the noise level of the input SEC-SAXS data, the degree of overlap of the concentration profiles of components and their asymmetry, which is important for its practical applications.

Results: By quantitative analysis of small angle scattering data using parametric linear and nonlinear minimisation methods, we were able to: (i) reveal the encounter-stochastic mechanism of adrenodoxin (Adx) complex formation with mitochondrial cytochrome c (Cc) [7]; (ii) show the ability of the enzyme lumazine synthase to self-assemble into icosahedral capsids ranging in size from 16 to 32 nm, and their composition depends on both the presence of protein mutations and changes in physicochemical environmental conditions [8]; (iii) analyze the evolution of the structural parameters of DMPC/DPPPC vesicles during extrusion, allowing the production of single-layer vesicles with a radius

of (30±1) nm [4]; (iv) for mixtures consisting of viral lipids and matrix protein M1 of influenza A virus, evaluate the effect of hemagglutinin on the formation of lipoprotein nanodomains in the lipid bilayer [9].

The particle shape reconstruction algorithm for multicomponent mixtures was successfully applied to study the formation of insulin protein fibrils and allowed us to determine the shape of the intermediate component in the process of their growth, which is an elongated structure with a size of 20 nm, consisting of five bound domains with sizes close to insulin monomers, which confirms the hypothesis of an oligomeric precursor, which is the primary elongation unit of insulin amyloid fibrils [5]. Using the proposed method, a three-dimensional shape of hetero-tetramer was reconstructed for the nerve growth factor protein (case of a two-component mixture) from a concentration seria of SAXS data [5].

Using the EFA algorithm it was determined that for the pyruvate aldolase protein mixture its oligomeric components in solution are hexamers and octamers [6]. The reconstructed scattering profiles of the components agree well with the scattering curves calculated from the known crystallographic models of proteins.

Conclusion: The developed methods significantly expand the possibilities of analyzing multicomponent nanoscale systems by the SAXS method and can be applied to a wide range of biological, organic and composite systems.

Funding: The work was carried out within the state assignment of the NRC “Kurchatov Institute”.

Список литературы/References

1. Konarev P.V., Volkov V.V., Sokolova A.V., Koch M.H.J., Svergun D.I. PRIMUS – a Windows-PC based system for small-angle scattering data analysis. *J Appl Cryst.* 2003;36:1277-1282
2. Manalastas-Cantos K., Konarev P.V., Hajizadeh N.R. et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Cryst.* 2021;54:343-355
3. Konarev P.V., Petoukhov M.V., Dadinova L.A., Fedorova N.V., Volynsky P.E., Svergun D.I., Batishchev O.V., Shtykova E.V. BILMIX: a new approach to restore the size polydispersity and electron density profiles of lipid bilayers from liposomes using small-angle X-ray scattering data. *J Appl Cryst.* 2020;53:236-243
4. Konarev P.V., Gruzinov A.Y., Mertens H.D.T., Svergun D.I. Restoring structural parameters of lipid mixtures from small-angle X-ray scattering data. *J Appl Cryst.* 2021;54:169-179
5. Konarev P.V., Svergun D.I. Direct shape determination of intermediates in evolving macromolecular solutions from small-angle scattering data. *IUCrJ.* 2018;5:402-409
6. Konarev P.V., Graewert M.A., Jeffries C.M., Fukuda M., Cheremnykh T.A., Volkov V.V., Svergun D.I. EFAMIX, a tool to decompose inline chromatography SAXS data from partially overlapping components *Protein Sci.* 2022;31:269-282
7. Xu X., Reinle W., Hannemann F., Konarev P.V., Svergun D.I., Bernhardt R., Ubbink M. Dynamics in a pure encounter complex of two proteins studied by solution scattering and paramagnetic NMR spectroscopy. *J Am Chem Soc.* 2008;130:6395-6403
8. Zhang X., Konarev P.V., Petoukhov M.V., Svergun D.I., Xing L., Cheng R.H., Haase I., Fischer M., Bacher A., Ladenstein R., Meining W. Multiple assembly states of lumazine synthase: a model relating catalytic function and molecular assembly. *J Mol Biol.* 2006;362:753-770
9. Kordyukova L.V., Konarev P.V., Fedorova N.V. et al. The cytoplasmic tail of influenza A virus hemagglutinin and membrane lipid composition change the mode of M1 protein association with the lipid bilayer. *Membranes*. 2021;11:772

doi 10.18699/bgrs2024-3.1-10

В поисках Nα-ацетилтрансфераз бактерии *Thermus thermophilus:* идентификация, структурный анализ и предсказание субстратной специфичности

Лаптева Ю.*, Кудряшов Т., Локтюшов Е., Трунилина М., Дерюшева Е., Быков В., Соколов А.

ФИЦ «Пущинский научный центр биологических исследований» РАН, Институт биологического приборостроения, Пущино, Россия

** yulia.s.lapteva@gmail.com*

Ключевые слова: N-ацетилтрансфераза; GNAT-домен; N-концевое ацетилирование; субстратная специфичность; *Thermus thermophilus*

Мотивация и цель: Nα-ацетилтрансферазы (Nα-AT) катализируют перенос ацетильной группы от ацетилкоэнзима А на α-аминогруппу широкого спектра субстратов (аминокислоты, белки, серотонин и др.) [1]. Это большой класс ферментов, характеризующихся консервативной трехмерной структурой, но низким процентом идентичности последовательностей, что затрудняет их корректную аннотацию в геномах и предсказание их субстратной специфичности. Nα-AT входят в состав суперсемейства GNAT-ацетилтрансфераз, которое характеризуется специфическим чередованием элементов вторичных структур, а именно β0-β1-α1-α2-β2-β3-β4-α3-β5-α4-β6-β7, и их уникальной трехмерной укладкой – так называемый GNAT-домен [2]. Ферменты семейства Nα-AT (как и лизин-AT (KAT)) содержат единичный GNAT-домен, в то время как ферменты других семейств AT, могут содержать несколько GNAT, а также другие домены. Центральная часть GNAT-домена содержит петлю, связывающую ацетил-Ко А, которая имеет консенсус аминокислот Q/R-X-X-G-G/A. Учитывая важность Nα-ацетилирования в контроле активности ферментов [3] и функционирования белков (в том числе и факторов патогенности) в клетках [4]; участие в биосинтезе и деградации антибиотиков [5]; применение Nα-AT в биотехнологии ацетилирования рекомбинантных белков [6], актуальным представляется их корректное предсказание и аннотация в геномах микроорганизмов. Цель данной работы заключается в использовании алгоритмов и методов биоинформатики для поиска Nα-AT в геноме бактерии *T. thermophilus* HB8, их структурно-функционального анализа с целью предсказания их субстратной специфичности.

Методы и алгоритмы: Поиск аминокислотных последовательностей множества предполагаемых ацетилтрансфераз в геноме бактерии *T. thermophilus* проводили в NCBI по запросу “acetyltransferase”. Выравнивание аминокислотных последовательностей белков выполняли с использованием программ Clustal Omega [7] и/или T-coffee [8]. Предсказание вторичной структуры белков проводили при помощи онлайн-сервера PsiPred [9]. Предсказание третичной структуры белков осуществляли в программе Alphafold2 [10]. Для выравнивания третичных структур белков использовали программу PyMol. Поиск белков-ортологов проводили при помощи сервера BLAST и DALI [11]. С целью анализа предполагаемой ферментативной активности гены некоторых ферментов были клонированы с использованием

методов генной инженерии. Получены очищенные препараты ферментов. Nα-ацетилтрансферазная активность была исследована в реакции *in vitro* ацетилирования модельного белка [12].
Результаты: Анализ генома бактерии *T. thermophilus* HB8 на предмет наличия аннотированных "Nα-acetyltransferase" или "N-terminal acetyltransferase" не дал результата. Однако это свидетельствует не об отсутствии Nα-AT в бактерии, а о сложности адекватной аннотации ферментов данного многочисленного семейства. Менее специфичный поиск по запросу "acetyltransferase" позволил выявить свыше десяти аннотированных «предполагаемых» AT. Однако это значительно меньше, чем в бактерии *E. coli* (29 AT) или *Streptomyces* (72 AT), что опять свидетельствует о сложной идентификации и корректной аннотации ферментов данного семейства. Для отбора среди множества AT именно Nα-AT мы провели множественное выравнивание аминокислотных последовательностей исследуемых ферментов с детально изученными белками-ортологами из других бактерий. Выравнивание позволило отобрать шесть ферментов, несущих только один GNAT-домен, в основном это сравнительно небольшие белки, размером от 150–205 аминокислот. Выравнивание трехмерных структур отобранных шести предполагаемых Nα-AT выявило характерную для ферментов данного семейства топологию GNAT-домена, в особенности расположение двух β-листов, формирующих активный центр. Однако для трех из шести предполагаемых Nα-AT чередование элементов вторичных структур не соответствует каноническому за счет дополнительных вставок/делеций α-спиралей и/или β-структур. Множественное выравнивание последовательностей отобранных шести Nα-AT позволило идентифицировать консенсус аминокислот, связывающих ацетил-Ко А. У *T. thermophilus* он имеет вид $Q/_{R}$-$G/_{R}$-X-G-$L/_{V}$-$G/_{I}$, где нижним индексом отмечены остатки, встречающиеся только в одной последовательности Nα-AT из шести. Нами были клонированы, получены очищенные препараты ряда ферментов и исследована их Nα-ацетилтрансферазная активность в условия *in vitro* на модельном белке.
Выводы: Результаты работы позволили выделить специфические критерии для поиска, предсказания и отбора Nα-AT среди множеств GNAT-AT. Полученные данные в дальнейшем могут лечь в основу создания алгоритма идентификации Nα-AT в геномах других организмов, что в настоящее время, несомненно, очень актуально для корректной аннотации ферментов семейства Nα-AT и предсказания их субстратной специфичности.
Финансирование: Исследование выполнено при финансовой поддержке Российского научного фонда, грант № 23-24-00478, https://rscf.ru/project/23-24-00478/.

In search for Nα-acetyltransferases from the bacterium *Thermus thermophilus*: identification, structural analysis and prediction of substrate specificity

Lapteva Yu.*, Kudryashov T., Loktyushov E., Trunilina M., Deryusheva E., Bykov V., Sokolov A.

Federal Research Center "Pushchino Scientific Center for Biological Research", RAS, Institute of Biological Instrumentation, Pushchino, Russia
** yulia.s.lapteva@gmail.com*

Key words: N-acetyltransferase; GNAT domain; N-terminal acetylation; substrate specificity; *Thermus thermophilus*

Motivation and Aim: Nα-acetyltransferases (Nα-AT) catalyze the transfer of an acetyl group from acetyl coenzyme A to the α-amino group of a wide range of substrates (amino acids, proteins, serotonin, etc.) [1]. This is a large class of enzymes characterized by a conserved three-dimensional structure but a low percentage of sequence identity, which makes it difficult to correctly annotate them in genomes and predict their substrate specificity. Nα-ATs are part of the GNAT acetyltransferase superfamily, which is characterized by a specific alternation of secondary structure elements, namely "β0-β1-α1-α2-β2-β3-β4-α3-β5-α4-β6-β7", and their unique three-dimensional folding, the so-called "GNAT domain" [2]. Enzymes of the Nα-AT family (like lysine AT (KAT)) contain a single GNAT domain, while enzymes of other AT families may contain multiple GNATs as well as other domains. The central part of the GNAT domain contains an acetyl-Co A binding loop, which has the amino acid consensus "Q/R-X-X-G-G/A". Considering the importance of Nα-acetylation in the control of enzyme activity [3] and the functioning of proteins (including pathogenicity factors) in cells [4]; participation in the biosynthesis and degradation of antibiotics [5]; the use of Nα-AT in the biotechnology of acetylation of recombinant proteins [6], their correct prediction and annotation in the genomes of microorganisms seems relevant. The purpose of this work is to use algorithms and bioinformatics methods to search for Nα-AT in the genome of the bacterium T. *thermophilus* HB8, their structural and functional analysis in order to predict their substrate specificity.

Methods and Algorithms: A search for amino acid sequences of multiple putative acetyltransferases in the genome of the bacterium Thermus thermophilus was carried out in NCBI using the query "acetyltransferase". Alignment of amino acid sequences of proteins was carried out using the Clustal Omega [7] and/or T-coffee [8] programs. Prediction of the secondary structure of proteins was carried out using the PsiPred online server [9]. Prediction of the tertiary structure of proteins was carried out using the Alphafold2 program [10]. The PyMol program was used to align protein tertiary structures. The search for orthologous proteins was carried out using the BLAST and DALI servers [11]. In order to analyze putative enzymatic activities, the genes for some enzymes have been cloned using genetic engineering techniques. Purified enzyme preparations were obtained. Nα-acetyltransferase activity was studied in the in vitro reaction of acetylation of a model protein [12].

Results: Analysis of the genome of the bacterium T. thermophilus HB8 for the presence of an annotated "Nα-acetyltransferase" or "N-terminal acetyltransferase" was inconclusive. However, this does not indicate the absence of Nα-AT in the bacterium, but rather the difficulty of adequately annotating the enzymes of this large family. A less specific search for "acetyltransferase" identified over ten annotated "putative" Abs. However, this is significantly less than in the bacteria *E. coli* (29 AT) or *Streptomyces* (72 AT), which again indicates the complex identification and correct annotation of enzymes of this family. To select Nα-AT among many ATs, we carried out multiple alignment of the amino acid sequences of the enzymes under study with detailed orthologous proteins from other bacteria. The alignment allowed us to select six enzymes that carry only one GNAT domain, mostly relatively small proteins ranging in size from 150-205 amino acids. Alignment of the three-dimensional structures of the selected six putative Nα-ATs revealed the GNAT domain topology characteristic of enzymes of this

family, in particular the arrangement of the two β-sheets forming the active site. However, for three of the six putative Nα-ATs, the alternation of secondary structure elements does not correspond to the canonical one due to additional insertions/deletions of α-helices and/or β-structures. Multiple sequence alignment of selected six Nα-ATs made it possible to identify a consensus of amino acids binding acetyl-Co A. In *T. thermophilus*, it has the form "Q/R-G/R-X-G-L/V-G/I", where the subscript denotes residues found only in one Nα-AT sequence out of six. We cloned and obtained purified preparations of a number of enzymes and studied their Nα-acetyltransferase activity under in vitro conditions on a model protein.

Conclusion: The results of the work made it possible to identify specific criteria for searching, predicting and selecting Nα-ATs among the sets of GNAT-ATs. The data obtained can subsequently form the basis for creating an algorithm for identifying Nα-ATs in the genomes of other organisms, which is currently, undoubtedly, very important for the correct annotation of enzymes of the Nα-AT family and prediction of their substrate specificity.

Funding: Research was funded by a grant from the Russian Science Foundation, No. 23-24-00478, https://rscf.ru/project/23-24-00478/.

Список литературы/References

1. Deng S., Marmorstein R. Protein N-Terminal Acetylation: Structural Basis, Mechanism, Versatility, and Regulation. *Trends Biochem Sci.* 2021;46:15-27
2. Abboud A., Bedoucha P., Byska J., Arnesen T., Reuter N. Dynamics-function relationship in the catalytic domains of N-terminal acetyltransferases. *Comput Struct Biotechnol J.* 2020;18:532-547
3. Parks A.R., Escalante-Semerena J.C. Modulation of the bacterial CobB sirtuin deacylase activity by N-terminal acetylation. *Proc Natl Acad Sci USA.* 2020;117:15895-15901
4. Permyakov S.E., Vologzhannikova A.A., Emelyanenko V.I., Knyazeva E.L., Kazakov A.S., Lapteva Y.S., Permyakova M.E., Zhadan A.P., Permyakov E.A. The impact of alpha-N-acetylation on structural and functional status of parvalbumin. *Cell Calcium.* 2012;52:366-376
5. Favrot L., Blanchard J.S., Vergnolle O. Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation. *Biochemistry.* 2016;23:989-1002
6. Esipov R.S., Makarov D.A., Stepanenko V.N., Miroshnikov A.I. Development of the intein-mediated method for production of recombinant thymosin beta4 from the acetylated in vivo fusion protein. *J Biotechnol.* 2016;228:73-81
7. Sievers F., Higgins D.G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 2018;27:135-145
8. Notredame C., Higgins D.G., Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 2000;302:205-217
9. McGuffin L.J., Bryson K., Jones D.T. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000;16:404-405
10. Jumper J., Evans R., Pritzel A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583-589
11. Holm L. DALI and the persistence of protein shape. *Protein Sci.* 2020;29:128-140
12. Lapteva Y.S., Vologzhannikova A.A., Sokolov A.S., Ismailov R.G., Uversky V.N., Permyakov S.E. In Vitro N-Terminal Acetylation of Bacterially Expressed Parvalbumins by N-Terminal Acetyltransferases from Escherichia coli. *Appl Biochem Biotechnol.* 2021;193:1365-1378

doi 10.18699/bgrs2024-3.1-11

Совершенствование репортерных свойств люциферазы Renilla с использованием методов молекулярного моделирования

Ларионова М.Д.[1]*, Иванисенко Н.В.[1, 2, 3], Иванисенко В.А.[1, 2], Буракова Л.П.[1], Наташин П.В.[1], Еремеева Е.В.[1], Высоцкий Е.С.[1]

[1] *Федеральный исследовательский центр «Красноярский научный центр СО РАН», Институт биофизики СО РАН, Красноярск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[3] *AIRI, Москва, Россия*

* *Larionova.marina@inbox.ru*

Ключевые слова: биолюминесценция; мутагенез; ProteinMPNN; AlfaFold2; целентеразин; рациональный дизайн белков; биолюминесцентный репортер

Люцифераза Renilla представляет собой односубъединичный белок семейства α/β-гидролаз, ответственный за биолюминесценцию морского мягкого коралла *Renilla*. Люцифераза дикого типа катализирует окислительное декарбоксилирование субстрата целентеразина с излучением голубого света (λ_{max} = 480 нм) [1]. Ранее многократно предпринимались попытки совершенствования люциферазы Renilla, которые включали мутагенез на основе консенсусных последовательностей α/β-гидролаз [2], случайный мутагенез [3, 4], а также внедрение дисульфидных связей [5]. Одним из самых перспективных и высокоактивных вариантов люциферазы Renilla является «желтый» мутант люциферазы (RmY), излучающий с максимумом 535 нм с нативным целентеразином (CTZ) и 574 нм в реакции с синтетическим аналогом CTZ-v [6]. «Желтый» мутант люциферазы зарекомендовал себя как удобный внутриклеточный сенсор для биолюминесцентного имиджинга.

Цель: Современные биоинформатические подходы позволяют не только модернизировать ферментативные функции белков, но и *de novo* предсказывать белки с заданными свойствами [7]. Поэтому рациональный дизайн с целью улучшения ферментативных характеристик и спектральных свойств RmY представляет большой интерес с точки зрения совершенствования RmY как репортерного биолюминесцентного белка.

Методы: Для предсказания аминокислотных замен, способных изменять физико-химические свойства люциферазы Renilla (RmY) использовались современные методы молекулярного моделирования и глубокого машинного обучения, включая модели обратного фолдинга (ProteinMPNN). Для этого были реконструированы модели Renilla в комплексе с CTZ/CTZ-v, на основе которых предсказаны мутации, способные изменять полярное окружение субстрата, увеличивать термостабильность фермента и влиять на кинетику биолюминесценции.

Результаты: Предсказанные точечные и множественные замены были внесены в последовательность RmY для получения мутантных вариантов люциферазы посредством экспрессии в *E. coli*. В результате первичного скрининга на бактериальных лизатах были идентифицированы мутанты, в большей степени влияющие на ферментативные и спектральные характеристики люциферазы.

В дальнейшем удачные мутанты более тщательно были охарактеризованы в виде чистых белковых препаратов. В ходе работы был оптимизирован протокол получения препаратов люциферазы из бактериальной цитоплазмы с сохранением до 80 % биолюминесцентной активности и продукцией до 30 мг/л. В результате рационального мутагенеза идентифицированы мутанты RmY, демонстрирующие ожидаемые эффекты – повышение термостабильности и изменение спектров биолюминесценции. Были выявлены мутанты, которые характеризовались сдвигом температурного оптимума фермента от 26 до 37 °C; показано, что за эти изменения ответственна одиночная замена K189Q. Кроме того, анализ мутантов позволил выявить аминокислотные остатки, выполняющие критическую роль в образовании эмиттера и, соответственно, в изменении спектральных характеристик мутантных люцифераз.

Выводы: Таким образом, нам удалось продемонстрировать высокий потенциал методов предсказания последовательности белков на основе ProteinMPNN и AlphaFold2, позволивших провести рациональный дизайн люциферазы Renilla. В результате работы были получены термостабильные мутанты люциферазы Renilla с оптимумом работы при условиях культивирования клеток млекопитающих (37 °C). Кроме этого, идентифицированы новые мутанты, расширяющие ряд целентеразин-зависимых люцифераз с эмиссией в области от 400 нм (фиолетовой) до 575 нм (желтой). Проведенные компьютерное моделирование и экспериментальная работа позволили выявить новых мутантов люциферазы Renilla, имеющих преимущества для их использования в качестве биолюминесцентных репортеров для различных аналитических задач.

Финансирование: Исследование выполнено при поддержке гранта Российского научного фонда № 22-14-00125.

Improving the reporter properties of Renilla luciferase by using molecular modeling methods

Larionova M.D.[1]*, Ivanisenko N.V.[1, 2, 3], Ivanisenko V.A.[1, 2], Burakova L.P.[1], Natashin P.V.[1], Eremeeva E.V.[1], Vysotski E.S.[1]

[1] *Federal Research Center "Krasnoyarsk Science Center SB RAS', Institute of Biophysicsm SB RAS, Krasnoyarsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *AIRI, Moscow, Russia*

* *Larionova.marina@inbox.ru*

Key words: bioluminescence; mutagenesis; ProteinMPNN; AlfaFold2; coelenterazine; protein rational design; bioluminescent reporter

Renilla luciferase is a monomeric protein which belongs to the α/β-hydrolase enzyme family, responsible for the bioluminescence of the marine soft coral *Renilla*. The wild-type luciferase catalyzes the oxidative decarboxylation of the substrate coelenterazine, this being accompanied by blue light emission (λ_{max} = 480 nm) [1]. Multiple attempts have been made to improve Renilla luciferase, including mutagenesis based on consensus sequences of α/β-hydrolases [2], random mutagenesis [3, 4], and the

introduction of disulfide bonds [5]. One of the most promising and highly active variants of Renilla luciferase is the “yellow” mutant luciferase (RmY), emitting light with a peak at 535 nm with native coelenterazine (CTZ) and at 574 nm in reaction with a synthetic CTZ-v analog [6]. The one has proven to be a convenient intracellular sensor for bioluminescent imaging.

Aim: Modern bioinformatics approaches allow not only modification of the protein enzymatic functions but the de novo prediction of proteins with desired properties as well [7]. Therefore, the rational design aimed at enhancing RmY as a reporter bioluminescent protein in terms of its enzymatic characteristics and spectral properties is objectively of great interest.

Methods: To predict amino acid substitutions capable of altering the biophysical properties of Renilla luciferase (RmY), we applied state-of-art molecular modeling and deep learning methods, including inverse folding models (ProteinMPNN). For this purpose, the models of Renilla in complex with CTZ/CTZ-v substrates were reconstructed, based on which the mutations capable of altering the substrate's polar environment, enhancing thereby the enzyme thermostability, and affecting bioluminescence kinetics were predicted.

Results: The predicted point and multiple mutations were introduced into the RmY sequence to obtain mutant luciferase variants through expression in *E. coli* cells. As a result of primary screening conducted on bacterial lysates, the mutants that mostly affected the enzymatic and spectral characteristics of luciferase were identified.

The perspective mutants were further characterized as pure protein preparations. An optimized protocol to obtain luciferase samples from bacterial cytoplasm was developed, preserving up to 80 % bioluminescent activity and yielding up to 30 mg/L. Rational mutagenesis identified the RmY mutants demonstrating the expected effects – increased thermostability and changes in bioluminescence spectra. The mutants with a shift in the enzyme's temperature optimum from 26 °C to 37 °C were identified and a single substitution K189Q was shown to be responsible for the shift. In addition, the analysis of mutants revealed amino acid residues playing a critical role in in the formation of the emitter and, consequently, in altering the spectral characteristics of mutant luciferases.

Conclusions: Taken together, we succeeded in demonstrating high potential of ProteinMPNN and AlphaFold2 computational algorithms for protein structure prediction which allowed for the rational design of Renilla luciferase. As a result of the study, thermostable mutants of Renilla luciferase with optimal performance under mammalian cell culture conditions (37 °C) were obtained. Moreover, the novel identified mutants expanded the row of coelenterazine-dependent luciferases with the emission ranging from 400 nm (purple) to 575 nm (yellow). Computer modeling and experimental work have allowed us to identify new Renilla luciferase mutants that have advantages for their use as bioluminescent reporters for various analytical applications.

Funding: The study was supported by the grant of the Russian Science Foundation No. 22-14-00125.

Список литературы/References

1. Titushin M.S., Markova S.V., Frank L.A. et al. Coelenterazine-binding protein of Renilla muelleri: cDNA cloning, overexpression, and characterization as a substrate of luciferase photochem. *Photobiol Sci.* 2008;(7):189-196
2. Loening A.M., Fenn T.D., Wu A.M., Gambhir S.S. Consensus guided mutagenesis of Renilla luciferase yields enhanced stability and light output. *Protein Eng Des Sel*. 2006;19(9):391-400
3. Loening A., Wu A., Gambhir S. Red-shifted Renilla reniformis luciferase variants for imaging in living subjects. *Nat Methods.* 2007;4:641-643

4. Woo J., von Arnim A.G. Mutational optimization of the coelenterazine-dependent luciferase from Renilla. *Plant Methods.* 2008;(4):23
5. Kahrani Z.F., Emamzadeh R., Nazari M., Mahdi Rasa S.M. Molecular basis of thermostability enhancement of Renilla luciferase at higher temperatures by insertion of a disulfide bridge into the structure. *BBA Proteins Proteom*. 2017;1865(2):252-259
6. Stepanyuk G.A., Unch J., Malikova N.P. et al. Coelenterazine-v ligated to Ca^{2+}-triggered coelenterazine-binding protein is a stable and efficient substrate of the red-shifted mutant of Renilla muelleri luciferase. *Anal Bioanal Chem*. 2010;398:1809-1817
7. Yeh A.H.W., Norn C., Kipnis Y. et al., De novo design of luciferases using deep learning. *Nature.* 2023;614:774-780

doi 10.18699/bgrs2024-3.1-12

От изучения свойств к рациональному дизайну конструкций на основе нуклеиновых кислот

Ломзов А.

Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
lomzov@niboch.nsc.ru

Ключевые слова: производные и аналоги нуклеиновых кислот; гибридизационные свойства; термодинамика; структура; молекулярная динамика; прогностические модели

Мотивация и цель: Олигонуклеотиды широко используются в различных областях фундаментальных и прикладных исследований, в биотехнологических и биомедицинских применениях. Изменение свойств олигонуклеотидов для улучшения их потребительских качеств при решении разнообразных задач осуществляют путем выбора специфических нуклеотидных последовательностей, а в случае, если этого оказывается недостаточно, в структуру олигомеров вводят химические модификации. Такие модификации научились вводить в разнообразные положения нуклеотидной цепи: в гетероциклическое основание, остаток сахара или остаток фосфорной кислоты. Более того, возможно полное замещение таких фрагментов нуклеотидов, например, рибозофосфатного остова целиком [1, 2]. Уже разработано большое число таких модификаций, однако запрос на новые производные нуклеиновых (НК) кислот все еще сохраняется в связи с рядом причин, к которым можно отнести патентную закрытость, сложность или неудобство синтеза, высокую цену синтеза и др. Данная работа посвящена разработке методов анализа физико-химических свойств олиго-нуклеотидов с модифицированным рибозофосфатным остовом, а также созданию подходов и алгоритмов для прогностического расчета таких свойств.
Методы и алгоритмы: При выполнении исследований использован широкий спектр физико-химических методов исследований. Для изучения гибридизационных свойств олигонуклеотидов применяли метод термической денатурации с оптической регистрацией сигнала. При исследовании структуры олигонуклеотидов и их комплексов использовали методы кругового дихроизма, флуоресцентной спектроскопии, ЯМР и ЭПР спектроскопии, метод задержки в геле, метод атомно-силовой микроскопии. Методы молекулярного моделирования применяли для изучения структуры и динамики биомолекулярных структур, а также для разработки подходов прогностического расчета гибридизационных свойств нуклеиновых кислот и их производных.
Результаты: Разработан ряд методов и подходов для изучения свойств производных нуклеиновых кислот. При разработке новых производных и аналогов НК зачастую сложно синтезировать протяженные цепи, содержащие модификации. Однако необходимо иметь инструменты изучения свойств таких соединений и оценки перспективности их применения. На примере глицин-морфолиновых олигонуклеотидов мы разработали подход для количественного определения их гибридизационных свойств. Метод позволяет количественно определять термодинамические параметры формирования комплексов производных с ДНК и/или РНК и, следовательно, перспективность дальнейшей

разработки таких соединений. Для этого необходимо использовать олигомеры, содержащие всего 5–6 мономерных звеньев или более [1]. Применимость данного метода была верифицирована и на ряде новых фосфорамидных бензоазольных производных олигонуклеотидов, которые мы недавно разработали [3, 4]. Исследование их свойств показало большие перспективы применения, в частности в области молекулярной диагностики [5]. В целом для фосфорамидных производных НК ведутся разработки их использования в области биомедицины. Для эффективного применения таких соединений нами проведено детальное исследование их свойств на примере фосфорилгуанидиновых олигонуклеотидов (ФГО), несущих N,N-диметилимидазолидин-2-иминовую модификацию. В деталях изучены гибридизационные свойства ФГО и структури их комплексов, в том числе методами структурного ЯМР и молекулярного моделирования [6–8]. На основании детального анализа их свойств разработан метод расчета гибридизационных свойств ФГО, содержащих любое число модификаций в различных положениях олигонуклеотидов при различной ионной силе раствора. Такие данные позволяют осуществлять рациональный дизайн структуры олигомеров, содержащих ФГ-группы, что может быть использовано в системах ПЦР диагностики.

Нами проведено исследование применимости ФГО и ФАО в качестве праймеров в системах аллель-специфической ПЦР, в том числе с использованием реальных образцов [4, 9]. Исследования показали большое преимущество использования фосфорамидных производных нуклеиновых кислот и способность их справиться с задачами, которые не могут решить нативные праймеры.

Несмотря на успех в разработках методов и полуэмпирических моделей для прогностического расчета свойств новых производных, их создание является чрезвычайно долгим, трудоемким и дорогостоящим. В связи с развитием компьютерных технологий, методов компьютерного моделирования и анализа мы попробовали использовать методы молекулярного моделирования для расчета гибридизационных свойств НК и их производных. Метод молекулярной динамики (МД) успешно применяют для исследования структуры и динамики комплексов НК, а полученные результаты оказываются очень близки к экспериментальным [10, 11]. Мы разработали подходы на основе МД моделирования для предсказания термодинамических параметров формирования комплексов нативных олиго-нуклеотидов. Их применение к производным и аналогам НК не дает такой же хорошей предсказательной способности [4, 9], однако позволяет оценить перспективность таких соединений с точки зрения структуры их комплексов с цепями НК и их гибридизационной способности.

Разработанные методы и подходы можно применять не только для дизайна коротких олигомеров, но и при построении супрамолекулярных конструкций. Нами проведено исследование формирования конкатемерных комплексов НК, а также формирования на их основе замкнутых кольцевых конструкций различной размерности и молекулярности. В рамках данных исследований обнаружена способность образования комплексов НК с различной геометрий и структурой парой или тройкой олигонуклеотидов. Установлены закономерности образования ДНК/ДНК, РНК/РНК и ДНК/РНК комплексов [12, 13]. Их формирование и динамика подтверждены набором физико-химических методов и походов, а для доказательства молекулярности таких комплексов разработан оригинальный подход [12, 13]. Такие комплексы имеют перспективы для применения в качестве

усилителей в системах гибридизационного анализа, в системах доставки терапевтических нуклеиновых кислот в клетки и при решении других задач биомедицины.
Выводы: Таким образом, разработаны инструменты для изучения свойств производных нуклеиновых кислот, методы и подходы для создания прогностических моделей расчета их гибридизационных свойств как на основании экспериментальных данных, так и с использованием методов молекулярного моделирования.
Финансирование: Исследование поддержано проектами РНФ (№ 23-14-00358 и № 18-14-00357), государственного задания ИХБФМ СО РАН № 121031300042-1) и РФФИ (№ 19-34-90127 и № 20-04-00719).

From the study of properties to the rational design of structures based on nucleic acids

Lomzov A.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
lomzov@niboch.nsc.ru

Key words: derivatives and analogues of nucleic acids; hybridization; thermodynamics; structure; molecular dynamics; predictive models

Motivation and Aim: Oligonucleotides are widely used in various areas of fundamental and applied research, in biotechnological and biomedical applications. Changing the properties of oligonucleotide to improve their consumer qualities in solving various tasks is carried out by selecting specific nucleotide sequences, and if this is not sufficient, chemical modifications are introduced into the oligomer structure. Such modifications learned to be introduced in various positions of the nucleotide chain: in the nucleobase, sugar or phosphate residue. Moreover, it is possible to completely replace such nucleotide fragments, for example, backbone [1, 2]. A large number of such modifications have already been developed, but the demand for new nucleic acid (NA) derivatives remains due to a number of reasons, including protection by patent law, the complexity or inconvenience of synthesis, the high cost of synthetics, etc. This work is devoted to the development of methods of analysis of the physico-chemical properties of oligonucleotide with modified backbone, as well as the developments of approaches and algorithms for prognostic calculation of such properties.
Methods and Algorithms: A wide range of physico-chemical methods of research were used in the study. The method of thermal denaturation with optical signal registration was used to analyze the hybridization properties of oligonucleotide. In the study of the structure of oligonucleotide and their complexes used methods of circular dichroism, fluorescent spectroscopy, NMR and EPR spectroscopy, the method of in gel electrophoresis, and atomic-force microscopy. Methods of molecular modeling have been used to study the structure and dynamics of biomolecular structures, as well as to develop approaches for the prognostic calculation of hybridization properties of nucleic acids and their derivatives.

Results: A number of methods and approaches have been developed to study the properties of nucleic acid derivatives. When developing new NA derivatives and analogues, it is often difficult to synthesize extended chains containing modifications. However, approaches must be available to study the properties of such compounds and to assess the prospects for their use. Using the example of glycine-morpholine oligonucleotide, we have developed a method to quantifying their hybridizing properties. The method allows to determine the thermodynamic parameters of the formation of complexes derivatives with DNA and/or RNA and, therefore, the prospect of further development of such compounds. To do this, oligomers containing only 5–6 or more monomer units [1] should be used. The applicability of this method has also been verified on a number of new phosphoramide benzoazole oligonucleotide derivatives (PABAO), which we have recently developed [3, 4]. The study of their properties has shown great prospects of application, in particular, in the field of molecular diagnostics [5]. In general, development is under way for phosphoramide derivatives in the field of biomedical applications. For the effective application of such compounds we have carried out a detailed examination of their properties on the example of phosphoryl guanidine oligonucleotide (PGO) carrying N,N-dimethylimidazolidine-2-amin modification(s). The hybridization properties of PGO and the structures of their complexes are studied in detail, including the methods of structural NMR and molecular modeling [6–8]. Based on a detailed analysis of their properties, a method has been developed to calculate the hybridization properties of PGOs containing any number of modifications in desired oligonucleotide positions at different ion strength of the solution. Such data enable rational design of the structure of oligomers containing PG groups, which can be used in PCR diagnostics.

We conducted a study of the applicability of PGO and PABAO as primers in allele-specific PCR systems, including using real samples [4, 9]. Studies have shown the great advantage of using phosphoramide derivatives of nucleic acids in PCR diagnostics, and their ability to cope with challenges that native primers cannot solve.

Despite the success in developing methods and semi-empirical models for the prognostic calculation of the properties of new NA derivatives, their developments is extremely long, time-consuming and costly. In connection with the development of computer technologies, methods of computer modeling and analysis, we have tried to use methods for molecular modeling to calculate the hybridization properties of NAs and their derivatives. The method of molecular dynamics (MD) is successfully applied to the study of structure and dynamics of NA complexes, and the results on the structure and the dynamic of such complexes are very close to experimental [10, 11]. We have developed approaches, based on MD modeling, to predict the thermodynamic parameters of the formation of native oligonucleotide complexes. Their application to derivatives and analogues of NA does not give the same good predictive ability [4, 9], but allows to assess the prospects of such compounds in terms of the structure of their complexes with NA chains and their hybridization ability.

The methods and approaches developed can be applied not only for the design of short oligomers, but also for the construction of supramolecular structures. We conducted a study of the formation of concatemeric complexes of NA, as well as formation on their basis of closed ring structures of different size and molecularity. Within the framework of these studies, the ability to form complexes of NA with different geometry and structure by the pair or triple oligonucleotides has been discovered. The similarity in patterns of the formation of DNA/DNA, RNA/RNA and DNA/RNA complexes [12, 13]

have been established. Their formation and dynamics are confirmed by a set of physico-chemical methods and processes, and an original approach has been developed to prove the molecularity of such complexes [12, 13]. Such complexes have prospects for use as amplifiers in hybridization analysis systems, in therapeutic nucleic acid cell delivery systems and in other biomedical applications.

Conclusion: Thus, tools for the study of properties of nucleic acid derivatives, methods and approaches for the creation of prognostic models for the calculation of their hybridization properties have been developed, both on the basis of experimental data and molecular modeling method.

Funding: The study is supported by projects RSF (No. 23-14-00358 and No. 18-14-00357) by the Russian state funded project for ICBFM SB RAS (No. 121031300042-1) and RFBR (No. 19-34-90127 и No. 20-04-00719).

Список литературы/References

1. Golyshev V.M., Abramova T.V., Pyshnyi D.V., Lomzov A.A. A new approach to precise thermodynamic characterization of hybridization properties of modified oligonucleotides: Comparative studies of deoxyribo- and glycine morpholine pentaadenines. *Biophys Chem*. 2018;234:24-33. doi 10.1016/j.bpc.2017.12.004
2. Golyshev V.M., Abramova T.V., Pyshnyi D.V., Lomzov A.A. Structure an Hybriization Properties of Glycine Morpholine Oligomers in Complexes with DNA an RNA: Experimental an Molecular Dynamics Stuies. *J Phys Chem B*. 2019;123(50):10571-10581. doi 10.1021/acs.jpcb.9b07148
3. Vasilyeva S.V., Baranovskaya E.E., Dyudeeva E.S., Lomzov A.A., Pyshnyi D.V. Synthesis of Oligonucleotides Carrying Inter-nucleotide N-(Benzoazole)-phosphoramide Moieties. *ACS Omega*. 2023;8(1):1556-1566. doi 10.1021/acsomega.2c07083
4. Vasilyeva S.V., Baranovskaya E.E., Chubarov A.S., Lomzov A.A., Pyshny D.V. Phosphoramide azole oligonucleotides, a method for the synthesis of phosphoramide azole oligonucleotides and a method for template enzymatic DNA synthesis using them. Application for Russ. patent, priority dated 04/26/23. Reg. No. 2023110694
5. Chubarov A.S., Baranovskaya E.E., Oscorbin I.P. et al. Phosphoramidate Azole Oligonucleotides for Single Nucleotide Polymorphism Detection by PCR. *Int J Mol Sci*. 2024;25(1):617. doi 10.3390/IJMS25010617
6. Lomzov A.A., Kupryushkin M.S., Dyudeeva E.S., Pyshnyi D.V. A Comparative Study of the Hybridization of Phosphoryl Guanidine Oligonucleotides with DNA and RNA. *Russ J Bioorganic Chem*. 2021;47(2):461-468. doi 10.1134/S1068162021020151
7. Lomzov A.A., Golyshev V.M., Dyudeeva E.S., Kupryushkin M.S., Pyshnyi D.V. Structure and hybridization properties of phosphorylguanidine oligonucleotides. *J Biomol Struct Dyn*. 2019;37:83-84
8. Golyshev V.M., Pyshnyi D.V., Lomzov A.A. Effects of Phosphoryl Guanidine Modification of Phosphate Residues on the Structure and Hybridization of Oligodeoxyribonucleotides. *J Phys Chem B*. 2021;125(11):2841-2855. doi 10.1021/acs.jpcb.0c10214
9. Chubarov A.S., Oscorbin I.P., Filipenko M.L., Lomzov A.A., Pyshnyi D.V. Allele-specific PCR for *KRAS* mutation detection using phosphoryl guanidine modified primers. *Diagnostics*. 2020;10(11):872. doi 10.3390/diagnostics10110872
10. Shevelev G.Y., Krumkacheva O.A., Lomzov A.A. et al. Physiological-temperature distance measurement in nucleic acid using triarylmethyl-based spin labels and pulsed dipolar EPR spectroscopy. *J Am Chem Soc*. 2014;136(28):9874-9877. doi 10.1021/ja505122n
11. Lomzov A.A., Sviridov E.A., Shernuykov A.V., Shevelev G.Y., Pyshnyi D.V., Bagryanskaya E.G. Study of a DNA duplex by nuclear magnetic resonance and molecular dynamics simulations. Validation of pulsed dipolar electron paramagnetic resonance distance measurements using triarylmethyl-based spin labels. *J Phys Chem B*. 2016;120(23):5125-5133. doi 10.1021/acs.jpcb.6b03193
12. Zamoskovtseva A.A., Golyshev V.M., Kizilova V.A., Shevelev G.Y., Pyshnyi D.V., Lomzov A.A. Pairing nanoarchitectonics of oligodeoxyribonucleotides with complex diversity: concatemers and self-limited complexes. *RSC Adv*. 2022;12(11):6416-6431. doi 10.1039/D2RA00155A
13. Kanarskaya M.A., Pyshnyi D.V., Lomzov A.A. Diversity of Self-Assembled RNA Complexes: From Nanoarchitecture to Nanomachines. *Molecules*. 2023;29(1):10. doi 10.3390/MOLECULES29010010/S1

doi 10.18699/bgrs2024-3.1-13

Анализ взаимодействий dCas9-белков с ДНК в *in vitro* системах для создания усовершенствованных методов управления работой геномов на основе CRISPR/(d)Cas-систем

Мамаева Н.Ю.[1]*, Глухов Г.С.[1, 2], Фескин П.Г.[1], Кристовский Н.В.[1], Шайтан А.К.[1]

[1] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

[2] *Университет МГУ-ППИ в Шэньчжэне, Шэньчжэнь, Китай*

* *mamaeva19n@gmail.com*

Ключевые слова: Cas-белки; CRISPR-(d)Cas системы; связывание с ДНК; генетические схемы; регуляция транскрипции; редактирование генома

Мотивация и цель: Система CRISPR/Cas9, широко используемая в качестве инструмента редактирования генома, может быть адаптирована для регуляции экспрессии генов путем активации или репрессии транскрипции, изменения эпигенома, визуализации динамики геномных локусов и построения генетических сетей [1]. Однако наличие неспецифического связывания является серьезной проблемой, ограничивающей применение системы [2]. В частности, рациональный дизайн генетических схем требует точных значений параметров взаимодействий комплексов dCas9 с целевыми сайтами ДНК. Целью работы было изучение взаимодействия dCas9-белков с ДНК *in vitro* в зависимости от ряда факторов (последовательность ДНК, РНК, состав растворителя) и разработка подходов к оптимизации связывания dCas9-белков с ДНК для создания функциональных генетических элементов.

Методы и алгоритмы: Эксперименты проводились на очищенном в ходе данной работы белке SpdCas9. Параметры взаимодействия комплексов вычислялись путем измерения поляризации флуоресценции в 384-луночных планшетах на планшетном сканере. Возможность сборки модифицированных комплексов проверяли методом просвечивающей электронной микроскопии с негативным контрастом. Для изучения конформационной подвижности apo-dCas9 использовали криоэлектронную микроскопию.

Результаты: Разработан метод характеризации параметров связывания на основе измерения поляризации флуоресценции, оптимизированный для изучения dCas9-белков и их комплексов с ДНК. На основе разработанного метода были измерены константы связывания комплексов dCas-гРНК с ДНК в условиях различных концентраций одно- и двухвалентных ионов. Так, показано, что присутствие ионов марганца в небольших концентрациях способствует улучшению сборки комплексов dCas9-гРНК с ДНК, поскольку вызывает локальное раскручивание ДНК в PAM-области. Однако повышение концентрации значительно ухудшает связывание, что объясняется изменением конформации ДНК, препятствующей связыванию с комплексом dCas9-гРНК. Также было охарактеризовано влияние последовательности ДНК на неспецифическое связывание данных комплексов и показана прямая зависимость от количества PAM-сайтов. В случае комплексов apo-dCas9 с ДНК показано, что взаимодействие носит электростатический

характер и зависит от соотношения концентрации положительных зарядов в PAM-взаимодействующем домене белка dCas9 и отрицательных зарядов ДНК. Так, при соотношении зарядов 1:1 происходит образование нерастворимых агрегатов, состоящих из dCas9 и ДНК, что важно учитывать в анализах *in vitro*. Также получена структура белка dCas9 с улучшенным разрешением в области PI домена. Анализ структуры, полученной методом CryoEM, показывает наличие подвижных доменов (RECIII), что согласуется с механизмом связывания белка. Однако наличие электронной плотности в области PI-домена говорит о его конформационной стабильности по сравнению с доменом RECIII. На основе литературных данных о нецелевом связывании предложен подход к повышению специфичности связывания путем введения дополнительного димеризационного домена в последовательность гРНК [3] и рассчитана эффективность сборки различных модифицированных конструкций *in vitro*.

Выводы: Применение и разработка различных методов как физико-химической, так и структурной характеризации системы CRISPR/Cas позволяют разобраться в механизмах нецелевого связывания Cas-белков с ДНК.

Финансирование: Исследование поддержано Министерством науки и высшего образования Российской Федерации (соглашение № 075-15-2024-539).

Modulation of dCas9-protein-DNA interactions by variation of solvent ionic composition and guide RNA functionalization

Mamaeva N.Y.[1]*, Glukhov G.S.[1, 2], Feskin P.G[1], Kristovskiy N.V.[1], Shaytan A.K.[1]

[1] *Lomonosov Moscow State University, Moscow, Russia*

[2] *Shenzhen MSU-BIT University, Shenzhen, China*

* *mamaeva19n@gmail.com*

Key words: Cas proteins; CRISPR-(d)Cas systems; DNA binding; genetic circuits; transcriptional regulation; genome editing

Motivation and Aim: The CRISPR/Cas9 system, widely used as a genome editing tool, can be adapted for gene expression regulation by activating or repressing transcription, modifying the epigenome, visualizing the dynamics of genomic loci, and constructing genetic circuits designing genetic circuits [1]. However, off-target activity and non-specific binding are significant issues that limit the application of the system. Specifically, the rational design of genetic circuits requires precise values of the interaction parameters of dCas9 complexes with target DNA sites [2]. The aim of this work was to study the interaction of dCas9 proteins with DNA *in vitro* depending on several factors (DNA sequence, RNA, solvent composition) and to develop approaches to optimize the binding of dCas9 proteins to DNA for the creation of functional genetic elements.

Methods and Algorithms: Experiments were conducted on the SpdCas9 protein purified during this work. Interaction parameters of the complexes were calculated by measuring fluorescence polarization in 384-well plates using a plate reader. The assembly of modified complexes was checked using negative-stain TEM. Cryo-EM was used to study the conformational mobility of apo-dCas9.

Results: A method for characterizing binding parameters based on fluorescence polarization measurements was developed and optimized for studying dCas9 proteins and their complexes with DNA. Using this method, the Kd of dCas-RNA complexes with DNA were measured under varying concentrations of mono- and divalent ions. It was shown that the presence of manganese ions in low concentrations facilitates the assembly of dCas9-RNA complexes with DNA by causing local unwinding of DNA in the PAM region. However, increasing the concentration significantly impairs binding due to DNA conformational changes that hinder interaction with the dCas9-RNA complex. Additionally, the influence of DNA sequence on the non-specific binding of these complexes was characterized, showing a direct dependence on the number of PAM sites. For apo-dCas9 complexes with DNA, it was shown that the interaction is electrostatic and depends on the ratio of positive charges in the PAM-interacting domain of the dCas9 protein to the negative charges of DNA. At a charge ratio of 1:1, insoluble aggregates consisting of dCas9 and DNA form, which is important to consider in *in vitro* analyses. The structure of the dCas9 protein with improved resolution in the PI domain was also obtained. Analysis of the structure obtained by CryoEM shows the presence of mobile domains (RECIII), which is consistent with the protein's binding mechanism. However, the presence of electron density in the PI domain region indicates its conformational stability compared to the RECIII domain.

Based on literature data on off-target binding, an approach to enhance binding specificity was proposed by introducing an additional dimerization domain into the RNA sequence, and the efficiency of assembling various modified constructs in vitro was calculated [3].

Conclusion: The application and development of various methods for both physicochemical and structural characterization of the CRISPR/Cas system allow for a better understanding of the mechanisms of off-target binding of Cas proteins to DNA.

Funding: The study is supported by the Russian Ministry of Science and Higher Education (grant No. 075-15-2024-539).

Список литературы/References

1. Fadul S.M., Arshad A., Mehmood R. CRISPR-based epigenome editing: mechanisms and applications. *Epigenomics*. 2023;15(21):1137-1155
2. Lazzarotto C.R. et al. CHANGE-seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity. *Nat Biotechnol*. 2020;38(11):1317-1327
3. Cho S.W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res*. 2014;24(1):132-141

doi 10.18699/bgrs2024-3.1-14

PepString: инструмент для поиска коротких пептидов

Мацуга Д.Г.[1]*, Суваан Б.С.[2], Козин С.А.[3], Анашкина А.А.[2, 3]

[1] *ФГБОУ ВО Астраханский ГМУ Минздрава России, Астрахань, Россия*

[2] *ФГАОУ ВО Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия*

[3] *ФГБУН Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

* *HitokiriDeepWeb@yandex.ru*

Ключевые слова: короткая последовательность; PepString; короткий пептид; пептидный поиск

Мотивация и цель: В последнее время появляется все больше работ по изучению пептидов различных размеров и их роли в регуляции различных процессов в организме, а также по поиску потенциальных сайтов связывания в белках. Короткие пептиды могут появляться в клетке несколькими путями, например, в процессе созревания белков или в результате деградации. В настоящее время на сайте https://www.uniprot.org нет возможности выполнять поиск по пептидам размером менее 7 аминокислот. Кроме того, Uniprot не позволяет выбор произвольного таксона для поиска, например, «млекопитающие», а только конечный таксон в цепочке (вид/штамм). Целью данной работы было создание веб-интерфейса для поиска происхождения коротких пептидов с возможностью их сортировки по таксономии.

Методы и алгоритмы: Для создания базы данных последовательностей белков на основе UniprotKB (SwissProt+TrEMBL) и ее автоматического обновления использовался Python и PostgreSQL; веб-интерфейс написан с помощью фреймворка Django.

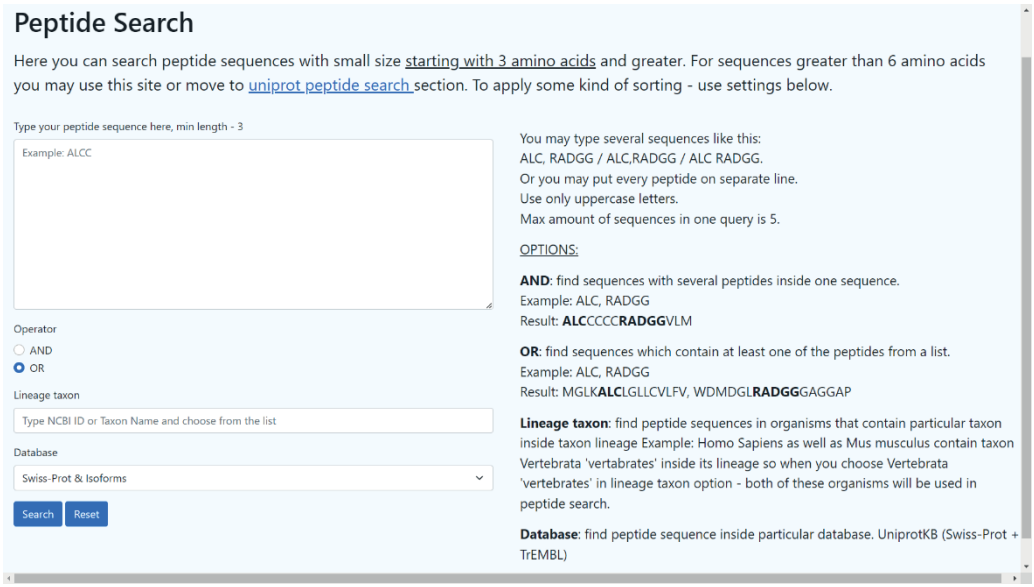


Рис. 2. Форма для запроса поиска коротких пептидов сервера PepString (http://pepstring.eimb.ru)

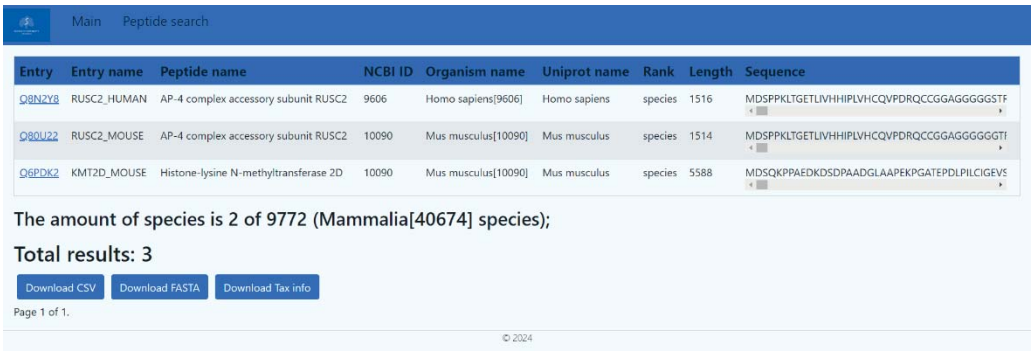


Рис. 3. Форма выдачи результатов запроса поиска коротких пептидов. Перечислены все белковые последовательности организмов выбранного таксона, содержащие подпоследовательность, совпадающую с запросом

Результаты: Создан сервер PepString (http://pepstring.eimb.ru), позволяющий искать короткие пептиды как в нужном таксоне, так и во всей базе данных, причем запрос может быть построен на основе наличия нескольких точно совпадающих фрагментов в одной последовательности (до пяти таких фрагментов), а также наличия хотя бы одного из перечисленных фрагментов в последовательности белка. Кроме того, можно искать последовательности как в UniprotKB, так и отдельно в SwissProt. Скриншот страницы запроса представлен на рис. 1. На рис. 2 приведена форма выдачи результатов запроса. Результаты пользователь может сохранить в формате FASTA, CSV, а также может получить подробную информацию о таксономии найденных последовательностей.
Выводы: Создан удобный инструмент для поиска коротких последовательностей пептидов с возможностью сортировки по таксономии.
Финансирование: Исследование поддержано грантом Российского научного фонда № 19-74-30007 на тему «Разработка новых фармакологических мишеней и взаимодействующих с ними низкомолекулярных химических соединений для лечения болезни Альцгеймера».

PepString: a tool for short peptide search

Matsuga D.G.[1]*, Suvaan B.S.[2], Kozin S.A.[3], Anashkina A.A.[2, 3]

[1] *FSBEI HE Astrakhan State Medical University of the Ministry of Health of Russia, Astrakhan, Russia*

[2] *FSAEI HE I.M. Sechenov First MSMU of the Ministry of Health of the Russian Federation (Sechenov University), Moscow, Russia*

[3] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

* *HitokiriDeepWeb@yandex.ru*

Key words: short sequence; PepString; short peptide; peptide search

Motivation and Aim: A lot of research works appear nowadays which explore various sized peptides and its role in different organism processes as well as research for finding potential binding sites in proteins. There are several ways for short peptides to appear in cell such as proteins maturation and degradation. Today there is no ability to search

sequences with size lesser than 7 amino acids on https://www.uniprot.org. Moreover, Uniprot taxonomy filtering doesn't provide ability to choose any taxon from lineage like mammalia. It only allows the last taxon in lineage (species/strain). The aim of this work was the creation of web interface for short protein origin search and ability for its taxonomy filtering.

Methods and Algorithms: Python and PostgreSQL were used to create protein sequence database on the base of UniprotKB (SwissProt+TrEmbl) and automatically update it; web interface was written using Django framework.

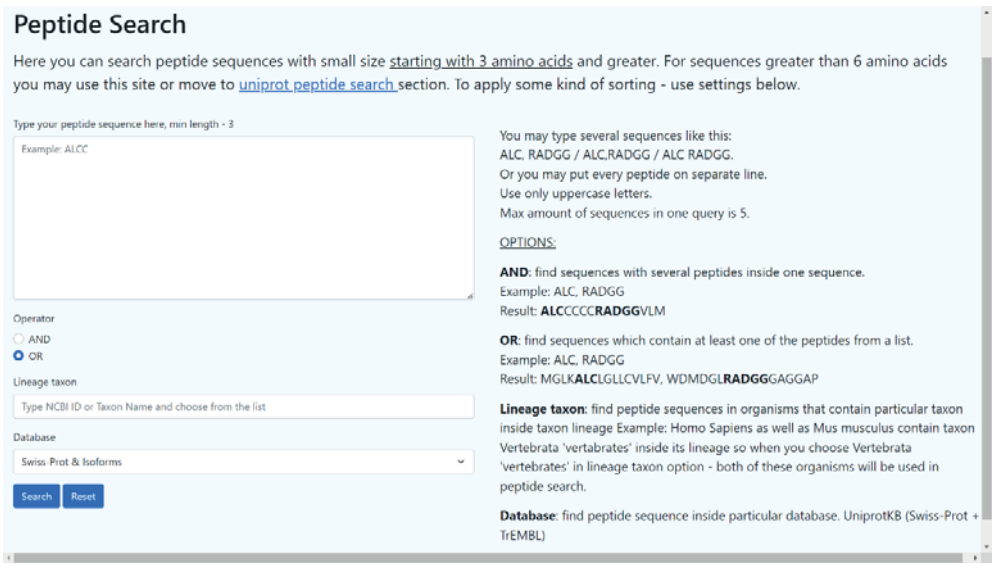


Fig. 4. Form that used to query for searching short peptides on the PepString server (http://pepstring.eimb.ru)

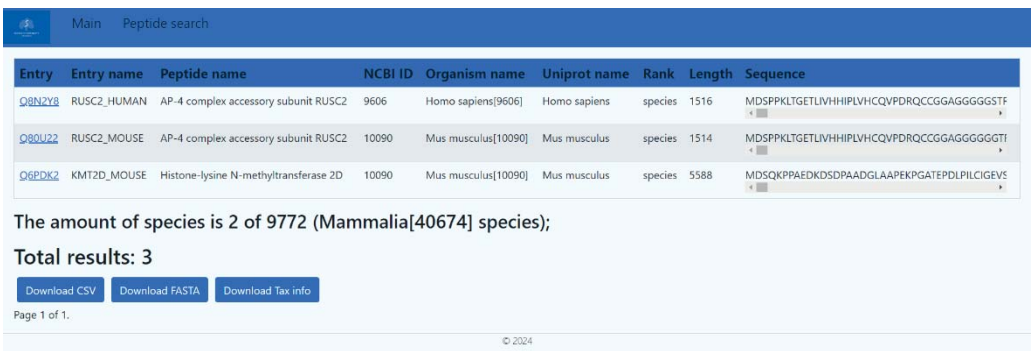


Fig. 5. Result table of short peptide search. All protein sequences of the chosen taxon which contain the needed subsequence are listed

Results: PepString server has been created (http://pepstring.eimb.ru) that allows user to search short peptides in a particular taxon or search it without filtering along the whole database. Also, query may be based on the presence of several fragments in the same sequence (five or less fragments) or on the presence of at least one fragment from a list in a sequence. In addition, there is a choice to search sequences in UniprotKB or in SwissProt. Page screenshot of the form is presented in the Fig. 1. In the Fig. 2 the result table is being shown. User also may save the result of query in FASTA and CSV formats and get thorough information about taxonomy of the resulting proteins.
Conclusion: A convenient tool for short peptide searching with taxonomy filtering was created.
Funding: The study is supported by grant from Russian science foundation No. 19-74-30007 on the topic “Development of new pharmacological targets and small molecular chemical compounds interacting with them for the treatment of Alzheimer’s disease”.

doi 10.18699/bgrs2024-3.1-15

Создание оптимальных структурных алфавитов для предсказания локальной и третичной структуры белка методами машинного обучения

Мильчевский Ю.В.[1]*, Мильчевская В.Ю.[1], Тевонян Л.Л.[1], Кретова А.Н.[1], Кравацкий Ю.В.[1, 2]

[1] *Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

[2] *Центр высокоточного редактирования и генетических технологий для биомедицины, Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

* *milch@eimb.ru*

Ключевые слова: предсказание структуры белков; локальная структура белка; структурные алфавиты (СА); выравнивание символьных последовательностей

Мотивация и цель: Предсказание структуры белка по аминокислотной последовательности является одной из центральных задач молекулярной биологии и биофизики. Несмотря на выдающиеся достижения методов машинного обучения (AlphaFold [1], ESMFold [2]), качество предсказания структуры белков, для которых данные о структуре гомологов отсутствуют в базах данных, по-прежнему часто остается низким. Более того, до сих пор нет общепризнанного представления локальной структуры [3–7]. Мы развиваем общие принципы формирования структурных алфавитов (СА) в зависимости от класса задач, для которых эти СА будут использоваться. Результатом применения этих принципов должно быть создание общего алгоритма формирования таких СА.

Методы и алгоритмы: Число структурных классов должно быть достаточным для адекватного представления хода главной цепи. Более длинные фрагменты требуют большего числа структурных классов для представления многообразия возможных конформаций белковой цепи. Ключевым параметром, определяющим выбор структурных элементов, является мера различия (метрика) между ними. В нашей работе [8] мы выбрали RMSD [9] и строго обосновали минимальное число структурных элементов заданной длины, необходимых для взаимно однозначного соответствия между координатами главной цепи и представлением локальной структуры, созданным на этих структурных элементах. В [8] приведен алгоритм, позволяющий однозначно восстановить 3D координаты по локальной структуре. Следовательно, для любой конформации фрагмента главной цепи по набору расстояний RMSD до каждого из структурных элементов можно восстановить его 3D координаты. Таким образом, мы **разложили многомерную задачу предсказания структуры белка на несколько** (по числу структурных классов) **отдельных одномерных задач**. Каждая такая задача направлена на количественную оценку взаимосвязи между последовательностью белка, статистикой его структурных конформаций и физико-химическими свойствами составляющих белок аминокислот. Для решения таких задач мы применяли пошаговый регрессионный анализ [10], который позволяет произвести отбор статистически значимых предикторов из большого первичного набора. Процедура выбора значимых предикторов позволяет проверять предположения о физико-

химических параметрах, определяющих структуру белка (физико-химические свойства аминокислот (>550) собраны в базе данных AAindex [11]). Для генерации предикторов нами предложена методика формирования и оптимизации наборов статистически значимых предикторов для последующего использования в задачах предсказания структуры и функций белка методами машинного обучения [12].

Восстановление декартовых координат по предсказанной локальной структуре, описанное в [8], на практике не работает ввиду быстрого накопления ошибок предсказания при сборке 3D структуры из отдельно предсказанных фрагментов. Для использования локальной структуры при сборке пространственной структуры всей белковой цепи мы предлагаем новый подход:

1. Предсказанная локальная структура записывается как символьная последовательность. Например, для PB [6] получается CA из 16 символов.
2. Для обучающей выборки предсказанные символьные последовательности CA, очевидно, однозначно выравниваются с символьными последовательностями, полученными из реальных трехмерных координат без вставок и делеций. Это позволяет создать по алгоритму [13] матрицу замен (substitution matrix), которая необходима для применения программ выравнивания [14].
3. При помощи программ выравнивания проводится сканирование банка данных PDB [15] для поиска «гомологичных» в смысле CA участков. Затем из таких участков с известной трехмерной структурой создаются начальные приближения для методов молекулярной динамики и механики.

Результаты: В нашей работе [16] мы применили нейронные сети множественной регрессии (multiple regression NN) для предсказания локальной и вторичной структур белка. Наш метод превзошел все методы, описанные в литературе, для классического набора негомологичных белковых цепей CB513 [17]: для вторичной структуры DSSP [7] Q3 = 85.99 %, Q8 = 79.35 %; для 16 PB [6] на выборке CB513 Q16 = 81.01 % [16]. Далее, для PB с использованием полученной в [16] предсказательной модели были выполнены предсказания локальных структур обучающей выборки, после преобразования которых в CA была построена матрица замен и оптимизирована процедура выравнивания методом Smith–Waterman из пакета FASTA [14].

В целях развития нашей методики был проведен расширенный поиск оптимальных структурных алфавитов и длин фрагментов, результатом чего стал CA, состоящий из 50 структурных элементов длиной 7 аминокислотных остатков. Для этого структурного алфавита были выполнены все этапы исследования, включая обучение нейронных сетей и создание аналогичной процедуры выравнивания.

Выводы: Процедура выравнивания предсказанных символьных последовательностей CA по базе символьных последовательностей, полученных из реальных трехмерных координат из PDB банка, не зависит от наличия гомологов в общепринятом смысле этого термина, поскольку наш алгоритм предсказания не использует информацию о наличии гомологии по аминокислотной последовательности, а базируется на физико-химических свойствах аминокислот и статистике встречаемости структурных элементов. Структурные алфавиты, использующие метрику RMSD, лучше (согласно нашему предварительному анализу), чем широко известные PB [6], описывают ход белковой цепи при длине структурных элементов от 5 до 11 остатков и позволяют

в итоге получить более адекватные начальные приближения для дальнейшего использования методов молекулярной динамики.
Финансирование: Исследование поддержано грантом Российского научного фонда РНФ (№ 24-24-00493).

Development of optimal structural alphabets for predicting protein local and tertiary structures by machine learning methods

Milchevskiy Y.V.[1]*, Milchevskaya V.Y.[1], Tevonyan L.L.[1], Kretova A.N.[1], Kravatsky Y.V.[1, 2]

[1] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

[2] *Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

* *milch@eimb.ru*

Key words: protein structure prediction; local protein structure; structural alphabets (SA); sequence alignment

Motivation and Objective: Protein structure prediction from sequence stands as a central task in molecular biology and biophysics. Despite remarkable advancements in machine learning methods (such as AlphaFold [1] and ESMFold [2]), the quality of protein structure prediction remains often low for proteins lacking structural homolog data in databases. Moreover, a universally accepted representation of local structure is still absent [3–7]. Here we develop general principles for constructing such representations, further referred to as Structural Alphabets (SAs), and tailoring them to specific tasks at hand.

Methods and Algorithms: The number of structural classes should be sufficient to adequately represent any protein backbone: namely, it should be possible to reconstruct the Cartesian coordinates of the backbone from its SA representation. A key parameter determining the selection of structural elements is the measure of dissimilarity (metric) between them. In [8], we have chosen RMSD [9] and rigorously justified the minimum number of structural elements of a given length necessary for a one-to-one correspondence between the backbone coordinates and the representation of local structure created from these structural elements. Further, we presented an algorithm to reconstruct the Cartesian coordinates from the SA representation.

Thus, we decomposed a highly complex problem of protein structure prediction into several one-dimensional problems, each corresponding to a single structural class. Each one-dimensional problem is aimed at quantitatively assessing the relationship between the protein sequence, the statistics of its structural conformations (dependent on the training data), and the physicochemical properties of the constituent amino acids (independent of the training data). To address these tasks, we employed stepwise regression analysis [10], which enables the selection of statistically significant predictors from a large initial set. The procedure for selecting significant predictors allows for testing suggestions about the physicochemical parameters that determine protein structure (physicochemical properties of amino acids (>550) are assembled in the AAindex database [11]), and thus increases interpretability of the method. We have

further proposed a methodology for generating and optimizing sets of statistically significant predictors for subsequent use in protein structure and function prediction [12]. Same as the exact SA classes, the predicted structural class assignments can be used to reconstruct the Cartesian coordinates of the backbone [8]. However, it is hindered by rapidly accumulating prediction error. To mitigate this, we propose a new approach: First, we represent a protein sequence in terms of its predicted SA assignment classes for each amino acid, which gives rise to a new symbolic sequence (for instance, Protein Blocks proposed in [6] give rise to a 16-symbol alphabet, and the structural class assignment corresponds to a fragment centered at the corresponding amino acid in the protein). Second, for the proposed symbolic representations, we perform a sequence alignment against symbolic representations of a large pool of proteins with known structures (for details see [13, 14]). While the alignment algorithms use symbolic sequences, there is structural information behind each of the SA symbols. This way, we identify structural homology between the query structure and the structures in the PDB [15], and use it to construct a "first 3D approximation" that serves as a starting point for molecular dynamics algorithms.

Results: We developed a method for local protein structure prediction using SAs, and evaluated it for the secondary structure prediction, DSSP Q8 class prediction and the 16-class Protein Block class prediction tasks using the CB513 data set [6, 7, 17]. In all tasks, it outperformed the existing methods showing the following results: Q3 = 85.99 %, Q8 = 79.35 %; Protein Blocks based Q16 = 81.01 % [16]. Further, we applied the 16 class Protein Blocks alphabet to fine-tune the alignment methods: namely, we used the true and predicted SA class assignments for the training set proteins to construct a substitution matrix for the aligner.

As another branch of this research, we aimed to identify an optimal set of SA classes and an optimal length of the corresponding peptides which resulted in a SA consisting of 50 structural elements with a length of 7 amino acids. For this structural alphabet, all stages of research were performed, including training the neural networks and creating a similar alignment procedure.

Conclusion: The proposed SA-based alignment procedure, as well as the local structure prediction algorithm, does not rely on the existence of sequence homologs in terms of the amino acid sequence. Instead, it relies on structural homologs, or homologs in terms of the SA-assigned sequence, which allows for structural prediction for proteins with no existing sequence homologs among the resolved structures – a protein class poorly covered even by the best available methods. For the query protein, it would be the predicted SA classes, and for the reference structures these are the true SA class assignments. Moreover, our method for the SA class prediction does not rely on amino acid sequence homology either but is based on physicochemical properties of amino acids and the occurrence statistics of the structural elements in the PDB. Using RMSD as a measure of dissimilarity, we observed that optimal length of a basic structural fragment to best describe the trajectory of the protein backbone is between 5 and 11. It ultimately provides a more adequate initial approximations for further application of molecular dynamics methods.

Funding: This research was supported by the Russian Science Foundation grant (No. 24-24-00493).

Список литературы/References

1. Jumper J. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583-589
2. Lin Z. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science.* 2023;379:1123-1130
3. Han K.F. et al. Three-dimensional structures and contexts associated with recurrent amino acid sequence patterns. *Protein Sci.* 1997;6:1587-1590
4. Sander O. et al. Local protein structure prediction using discriminative models. *BMC Bioinformatics.* 2006;7:14
5. Kloczkowski A. et al. Distance matrix-based approach to protein structure prediction. *J Struct Funct Genomics.* 2009;10:67-81
6. de Brevern A.G. et al. Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. *Proteins.* 2000;41:271-287
7. Kabsch W., Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 1983;22:2577-2637
8. Milchevskaya V. et al. Structural coordinates: A novel approach to predict protein backbone conformation. *PLoS One.* 2021;16:e0239793
9. Kabsch W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr, Sect A: Found Crystallogr.* 1976;32:922-923
10. Ralston A., Wilf H.S., Enslein K. Mathematical Methods for Digital Computers. New York: Wiley, 1960
11. Kawashima S. et al. AAindex: Amino acid index database, progress report 2008. *Nucleic Acids Res.* 2008;36:D202-D205
12. Milchevskiy Y.V. et al. Method to generate complex predictive features for machine learning-based prediction of the local structure and functions of proteins. *Mol Biol.* 2023;57:136-145
13. Henikoff J.G., Henikoff S. Using substitution probabilities to improve position-specific scoring matrices. *Comput Appl Biosci.* 1996;12:135-143
14. Pearson W.R. Blast and fasta similarity searching for multiple sequence alignment. *Methods Mol Biol.* 2014;1079:75-101
15. Berman H.M. et al. The Protein Data Bank at 40: Reflecting on the past to prepare for the future. *Structure.* 2012;20:391-396
16. Milchevskiy Y.V. et al. Effective local and secondary protein structure prediction by combining a neural network-based approach with extensive feature design and selection without reliance on evolutionary information. *Int J Mol Sci.* 2023;24(21):15656
17. Cuff J.A., Barton G.J. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins.* 1999;34:508-519

doi 10.18699/bgrs2024-3.1-16

Подбор условий кристаллизации химозинов жвачных животных

Миронова Е.[1, 2]*, Колыбалов Д.[1], Архипов С.[1], Дюсенова С.[1], Шевцов М.[3], Беленькая С.[4], Волосникова Е.[4], Борщевский В.[3], Щербаков Д.[4]

[1] *Центр коллективного пользования «Сибирский кольцевой источник фотонов» Института катализа и адсорбции СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Московский физико-технический институт, Долгопрудный, Россия*

[4] *ФБУН ГНЦ ВБ «Вектор», Новосибирск, Россия*

* *mironovaforwork@gmail.com*

Ключевые слова: кристаллизация; белок; химозин; рентгеноструктурный анализ

Мотивация и цель: Химозин представляет собой белок класса пептидаз, который активно используется в пищевой промышленности. Ранее для производства сыра за счет селективного процесса створаживания молока использовали химозин, содержащийся в измельченных сычугах юных телят, однако в настоящее время основной объем этого фермента получают биотехнологическим способом. Для химозинов различных видов наблюдается разная специфичность к каппа-казеину коровьего молока, в случае химозина верблюда превышающая даже показатели родного фермента [1]. Целью нашей работы является поиск условий кристаллизации ряда химозинов различных жвачных (алтайский марал, северный олень, архар) для их дальнейшего рентгеноструктурного анализа. Полученные данные позволят уточнить предполагаемый механизм действия фермента, а также провести анализ факторов, влияющих на субстратную специфичность.

Методы и алгоритмы: Кристаллизация химозинов алтайского марала, лося, архара и северного оленя проводилась методом сидячей капли. Для формирования капель использовался робот NT8 Drop Setter (Formulatrix, США). В качестве растворов осадителей были выбраны коммерчески доступных скрининговые наборы для кристаллизации: Pact Premier, Crystal Screen, Index, SG1, JCSG+ (Molecular Dimensions, Великобритания). Процедура скрининга включала в себя также варьирование соотношения объема раствора белка и осадителя. Для химозина алтайского марала выполняли рентгеноструктурный анализ с использованием синхротронного излучения. Съемку проводили на станции синхротронного источника излучения ID23-1 ESRF (Гренобль, Франция).

Результаты: Подобраны условия кристаллизации химозина марала, лося и архара. Кристаллы химозина марала были получены в следующих условиях: 0.2 M $CaCl_2$, 0.1 M Tris (pH 8.0), 20 % w/v PEG 6000; химозина лося – 0.2 M $CaCl_2$, 0.1 M ацетат натрия (pH 5.0), 20 % (w/v) PEG 6000 [1]; химозина архара – 0.01 M $ZnSO_4$, 0.1 M MES (pH 6.5), 25 % v/v PEG 550; 0.05 M ацетат цинка, 20 % w/v PEG 3350. Подобрать условия кристаллизации химозина северного оленя на данный момент не получилось. Возможно, это связано с малой концентрацией раствора исходного белка. Был выполнен рентгеноструктурный анализ полученных кристаллов в Европейском центре синхротронного излучения и в Шанхайском центре синхротронного излучения. Для химозина марала была расшифрована

кристаллическая структура и задепонирована в PDB (PDB ID: 8CIK). В случае химозина лося удалось определить только параметры элементарной ячейки. Для химозина архара в настоящее время ведутся работы по уточнению структурной модели.
Выводы: Были подобраны условия кристаллизации химозинов марала, лося и архара. Для химозина северного оленя требуется дальнейший поиск условий кристаллизации. Есть предположение, что оптимизация процедуры криопротектирования кристаллов химозина лося позволит получить от выращенных кристаллов набор дифракционных данных, с помощью которого можно будет выполнить решение и уточнение соответствующей структуры. Была расшифрована и задепонирована структура химозина марала.
Финансирование: Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации для ЦКП «СКИФ» ИК СО РАН (FWUR-2024-0040).
Шевцов М. и Борщевский В. выражают благодарность за поддержку Министерству науки и высшего образования Российской Федерации (соглашение № 075-03-2024-117, проект FSMG-2024-0012).

Screening for crystallization conditions of ruminant's chymosins

Mironova E.[1, 2]*, Kolybalov D.[1], Arkhipov S.[1], Diusenova S.[1], Shevtsov M.[3], Belenkaya S.[4], Volosnikova E.[4], Borshchevskiy V.[3], Shcherbakov D.[4]
[1] *Synchrotron Radiation Facility – Siberian Circular Photon Source "SKlF" Boreskov Institute of Catalysis, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Moscow Institute of Physics and Technology, Moscow, Russia*
[4] *The State Research Center of Virology and Biotechnology "Vector", Novosibirsk, Russia*
* *mironovaforwork@gmail.com*

Key words: crystallization; protein; chymosin; X-ray diffraction analysis

Motivation and Aim: Chymosin is a protein of the peptidase class, which is actively used in the food industry. Previously, chymosin contained in the crushed rennet of young calves was used for the production of cheese due to the selective process of milk coagulation, but currently the bulk of this enzyme is obtained by a biotechnological method. For different types of chymosins, different specificity to cow's milk kappa casein is observed, in the case of camel chymosin, even exceeding the parameters of the native enzyme [1]. The purpose of this work is to search for conditions for the crystallization of a number of chymosins of various ruminants (Altai maral, reindeer, argali) for their further X-ray diffraction analysis. The obtained data will clarify the proposed mechanism of action of the enzyme, as well as analyze the factors affecting substrate specificity.
Methods and Algorithms: Crystallization of chymosins of Altai maral, elk, argali and reindeer was carried out by the sitting-drop method. The NT8 Drop Setter robot (Formulatrix, USA) was used to form droplets. Commercially available screening kits for crystallization were selected as precipitant solutions: Pact Premier, Crystal Screen, Index, SG1, JCSG+ (Molecular Dimensions, England). The screening procedure also

included varying the volume ratio of the protein solution and precipitant. X-ray diffraction analysis using synchrotron radiation was performed for the Altai maral chymosin. The experiment was carried out at the station of the synchrotron radiation source ID 23-1 ESRF (Grenoble, France).

Results: Conditions of crystallization of chymosin of maral, elk and argali were selected. Maral chymosin crystals were obtained under the following conditions: 0.2 M $CaCl_2$, 0.1 M Tris (pH 8.0), 20 % w/v PEG 6000; moose chymosin – 0.2 M $CaCl_2$, 0.1 M sodium acetate (pH 5.0), 20 % (w/v) PEG 6000 [1]; argali chymosin – 0.01 M $ZnSO_4$, 0.1 M MES (pH 6.5), 25 % v/v PEG 550; 0.05 M zinc acetate, 20 % w/v PEG 3350. It has not been possible to select the conditions for crystallization of reindeer chymosin at the moment, perhaps due to the low concentration of the initial protein solution. X-ray diffraction analysis of the obtained crystals was performed at the European Synchrotron Radiation Center and at the Shanghai Synchrotron Radiation Center. The crystal structure of the maral chymosin was deciphered and deposited in PDB (PDB ID: 8 CIK). In the case of elk chymosin, it was possible to determine only the parameters of the unit cell. For argali chymosin, work is currently underway to refine the structural model.

Conclusion: The conditions of crystallization of the chymosins of maral, elk and argali were selected. Further search for crystallization conditions is required for reindeer chymosin. It is assumed that optimization of the cryoprotection procedure for moose chymosin crystals will allow obtaining a set of diffraction data from grown crystals, using which it will be possible to solve and refine the corresponding structure. The structure of the maral chymosin was deciphered and deposited.

Funding: This work was performed within the framework of a budget project of the Ministry of Science and Higher Education of the Russian Federation for Synchrotron radiation facility SKIF, Boreskov Institute of Catalysis (FWUR-2024-0040).

Shevtsov M. and Borshchevskiy V. express gratitude for the support of the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-03-2024-117, project FSMG-2024-0012).

Список литературы/References

1. Balabova D.V., Belash E.A., Belenkaya S.V., Shcherbakov D.N., Belov A.N. et al. Biochemical Properties of a Promising Milk-Clotting Enzyme, Moose (Alces alces) Recombinant Chymosin. *Foods.* 2023;12(20):3772. doi 10.3390/foods12203772

doi 10.18699/bgrs2024-3.1-17

Формирование белок-липидных доменов в плазматических мембранах

Мокрушников П.[1]*, Рудяк В.[1, 2]

[1] *Новосибирский государственный архитектурно-строительный университет (Сибстрин), Новосибирск, Россия*

[2] *Институт теплофизики им. С.С. Кутателадзе СО РАН, Новосибирск, Россия*

* *pavel.mokrushnikov@bk.ru*

Ключевые слова: белок-белковые и белок-липидные взаимодействиях; белок-липидные домены; β-структуры; структурные изменения плазматических мембран

Мотивация и цель: В настоящее время установлено, что при взаимодействии гормонов стресса и андрогенов с плазматическими мембранами происходит изменение их структуры [1, 2]. Под изменением структуры (конформации) плазматических мембран понимается изменение вторичной, третичной и четвертичной структур мембранных белков, фаз липидного бислоя, перераспределение белков и липидов по бислою, изменение их морфологии мембран. При этом в плазмалеммах меняются взаимодействия между белками и липидами, образуется неподвижная квазипериодическая сеть белок-липидных доменов, связанных с цитоскелетом. В этих доменах прилегающие к белкам липиды сильнее притягиваются к белкам, поменявшим свою конформацию, что ведет к изменению поля механических напряжений в мембране [3], функции мембран и клеток [4, 5]. Инициаторами образования белок-липидных доменов являются мембранные белки, которые изменили свою конформацию при взаимодействии мембраны с гормонами [1]. Вместе с тем все еще не ясно, какие именно изменения во вторичной структуре мембранных белков способствуют образованию вокруг них белок-липидных доменов. Целью данной работы является установление вторичных структур мембранных белков, появление которых способствует увеличению притяжения липидов к мембранным белкам и, как следствие, появлению белок-липидных доменов в мембране.

Методы и алгоритмы: В представленном докладе экспериментально исследуется изменение структуры мембран эритроцитов крыс при их связывании с андрогенами (андростероном, тестостероном, дегидроэпиандростероном (ДЭА), дегидроэпиандростерон сульфатом (ДЭАС)) и гормонами стресса (кортизолом, адреналином, норадреналином). С этой целью использовались флуоресцентные методы измерения тушения собственной флуоресценции мембранных белков, измерения микровязкости мембран с помощью зонда пирена и поглощения мембранами инфракрасного излучения.

Результаты: Методом измерения тушения собственной флуоресценции мембранных белков установлено, что все исследуемые гормоны связывались с мембранными белками. Максимальная величина тушения оказалось различной для разных гормонов. Это свидетельствует о разном сродстве гормонов с мембранными белками. Измерение микровязкости мембран показало, что при их связывании с гормонами, за исключением ДЭА, происходило образование неподвижной квазипериодической сети белок-липидных доменов, связанной с

цитоскелетом через узловой и анкириновый комплексы. При этом микровязкость липидного бислоя в белок-липидной области взаимодействия увеличивалась при меньшей удельной концентрации гормонов во взвеси. Это увеличение оказалось заметно больше, чем в липид-липидной области взаимодействия. Такое поведение свидетельствует о том, что усиливается взаимодействие между мембранными белками и окружающими их липидами. Эти липиды сместились ближе к мембранным белкам, произошло образование белок-липидных доменов. Липиды в этих доменах находятся в жидко упорядоченном состоянии L_o, липиды в пространстве между доменами находятся в жидко неупорядоченном состоянии L_d. При взаимодействии ДЭА с мембранами изменение микровязкости происходило одинаково в белок-липидной и липид-липидной областях взаимодействия. Это означает, что белок-липидные домены не сформировались, микровязкость увеличивалась равномерно по всему липидному бислою.

Причину такого разного взаимодействия гормонов с мембранами удалось установить с помощью измерения поглощения инфракрасного излучения. Оказалось, что по сравнению с образцами мембран эритроцитов, в которые добавлялись гормоны, в спектрах поглощения мембран, инкубированных с ДЭА, не появляются полосы поглощения 1685 $см^{-1}$, указывающие на появление в мембранных белках межмолекулярных β-листов [6]. Кроме того, не появляется полоса поглощения 1636 $см^{-1}$, присущая появлению в мембранных белках внутримолекулярных β-листов. При взаимодействии других гормонов с мембранами эти полосы в спектре поглощения мембран фиксировались, что говорит о переходе клубок→ β-структура в белках этих мембран. Вместе с тем при инкубации ДЭА с мембранами в их белках происходил переход клубок→α-структура. Об этом свидетельствует увеличение интегральной интенсивности полосы 1654 $см^{-1}$, которая является характерной полосой α-спиралей в пептидах. Таким образом, показано, что за образование белок-липидных доменов в плазматических мембранах ответственны не α-спирали, а β-структуры в мембранных белках.

Выводы: Установлено *in vitro*, что инициаторами появления белок-липидных доменов в плазматических мембранах, образующих неподвижную квазипериодическую сеть, связанную с цитоскелетом, являются мембранные белки, в которых после взаимодействия с гормонами появляются β-структуры. При этом появление новых α-спиралей в мембранных белках не усиливает взаимодействие между белками и окружающими их липидами, липиды не притягиваются ближе к белкам и не образуют белок-липидные домены.

Финансирование: Исследование поддержано мегагрантом Министерства науки и высшего образования РФ (Соглашение № 075-15-2021-575).

Formation of protein-lipid domains in plasma membranes

Mokrushnikov P.[1]*, Rudyak V.[1, 2]

[1] *Novosibirsk State University of Architecture and Civil Engineering (SIBSTRIN), Novosibirsk, Russia*

[2] *Institute of Thermophysics, SB RAS, Novosibirsk, Russia*

* *pavel.mokrushnikov@bk.ru*

Key words: protein-protein and protein-lipid interactions; protein-lipid domains; β-structures; structural changes in plasma membranes

Motivation and Aim: It has now been established that when stress hormones and androgens interact with plasma membranes, a membrane structural changes occur [1, 2]. Structural changes of plasma membranes mean changes in the secondary, tertiary and quaternary structures of membrane proteins, phases of the lipid bilayer, redistribution of proteins and lipids throughout the bilayer, and changes in membrane morphology. At the same time, the interactions between proteins and lipids in the plasma membranes change, and a stationary quasiperiodic network of protein-lipid domains associated with the cytoskeleton is formed. In these domains, lipids adjacent to proteins are more strongly attracted to proteins that have changed their conformation, which leads to a change in the mechanical stress field in the membrane [3], the function of membranes and cells [4, 5]. The initiators of the formation of protein-lipid domains are membrane proteins that change their conformation during the interaction of the membrane and hormones [1]. At the same time, it is still unclear what changes in the secondary structure of membrane proteins contribute to the formation of protein-lipid domains around them. The purpose of this work is to establish the secondary structures of membrane proteins, the appearance of which in membrane proteins contributes to an increase in the attraction of lipids to membrane proteins and, as a consequence, the appearance of protein-lipid domains in the membrane.

Methods and Algorithms: The rat erythrocyte membranes changes are studied experimentally at the presented report when they were bound to androgens (androsterone, testosterone, dehydroepiandrosterone (DHEA), dehydroepiandrosterone sulfate (DHEA-S)) and stress hormones (cortisol, adrenaline, norepinephrine). For this purpose, fluorescent methods were used to measure the quenching of the membrane proteins fluorescence intrinsic, to measure the microviscosity of membranes using a pyrene probe and the infrared radiation absorption.

Results: By measuring the membrane proteins fluorescence intrinsic quenching, it was established that all the studied hormones were bound to membrane proteins. The maximum magnitude of quenching was different for different hormones. This indicates a different affinity of hormones for membrane proteins. Membrane microviscosity measurements showed that when they were bound to hormones, with the exception of DHEA, a stationary quasiperiodic network of protein-lipid domains was formed, connected to the cytoskeleton through the junctional and ankyrin complexes. In this case, the microviscosity of the lipid bilayer in the protein-lipid interaction region began to increase at a lower specific concentration of hormones in the suspension. This increase was noticeably greater than in the lipid-lipid interaction region. This means that the interaction between membrane proteins and their surrounding lipids has increased. These lipids moved closer to the membrane proteins, and the formation of protein-lipid domains occurred. The lipids in these domains are in the liquid ordered state L_o, and the lipids in the space between the domains are in the liquid disordered state L_d. When DHEA interacted with membranes, the change in microviscosity occurred equally in the protein-lipid and lipid-lipid interaction regions. This means that protein-lipid domains were not formed, and the increase in microviscosity occurred uniformly throughout the lipid bilayer.

The reason for such different interactions of hormones with membranes was determined by measuring the absorption of infrared radiation. It turned out that, in comparison with

samples of erythrocyte membranes to which hormones were added, the absorption spectra of membranes incubated with DHEA did not contain absorption bands at 1685 cm^{-1}, indicating the appearance of intermolecular β-sheets in membrane proteins [6]. In addition, the absorption band at 1636 cm^{-1}, characteristic of the appearance of intramolecular β-sheets in membrane proteins, does not appear. When other hormones interacted with membranes, these bands appeared in the absorption spectrum of the membranes, which indicates a tangle → β-structure transition in the proteins of these membranes. At the same time, when DHEA was incubated with membranes, a tangle → α-helix transition occurred in their proteins. This is evidenced by an increase in the integral intensity of the band at 1654 cm^{-1}, which is a characteristic band of α-helix in peptides. Thus, it has been shown that it is not α-helix, but β-structures in membrane proteins that are responsible for the formation of protein-lipid domains in plasma membranes.

Conclusion: It has been established *in vitro* that the initiators of the protein-lipid domains in plasma membranes appearance, forming a stationary quasiperiodic network associated with the cytoskeleton, are membrane proteins in which β-structures appear after interaction with hormones. At the same time, the appearance of new α-helix in membrane proteins does not enhance the interaction between proteins and the surrounding lipids; lipids are not attracted closer to proteins and do not form protein-lipid domains.

Funding: The study is supported by a Megagrant from the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-575).

Список литературы/References

1. Мокрушников П.В., Панин Л.Е., Панин В.Е., Козельская А.И., Зайцев Б.Н. Структурные переходы в мембранах эритроцитов (экспериментальные и теоретические модели). Новосибирск: Новосибирский государственный архитектурно-строительный университет (Сибстрин), 2019
   [Mokrushnikov P.V., Panin L.E., Panin V.E., Kozelskaya A.I., Zaytsev B.N. Structural transitions in erythrocyte membranes (experimental and theoretical models). Novosibirsk: Novosibirsk State University of Architecture and Civil Engineering (Sibstrin), 2019 (in Russian)]
2. Mokrushnikov P.V. Mechanical Stresses in the Lipid Bilayer of Erythrocyte Membranes. In: Ashrafuzzaman M. (Ed.). Lipid Bilayers: Properties, Behavior and Interactions. NY: Nova Science Publishers, 2019;43-91
3. Мокрушников П.В., Рудяк В.Я. Механическая модель образования складок на плазматической мембране. *Доклады Академии наук высшей школы Российской Федерации*. 2023;2(59):29-40. doi 10.17212/1727-2769-2023-2-29-40
   [Mokrushnikov P.V., Rudyak V.Ya. Model of plasma membrane folds formation. *Doklady Akademii Nauk Vysshei Shkoly Rossiiskoi Federatsii = Proceedings of the Russian Higher School Academy of Sciences.* 2023;59(2):29-40. doi 10.17212/1727-2769-2023-2-29-40 (in Russian)]
4. Mokrushnikov P.V., Rudyak V.Ya., Lezhnev E.V. Mechanism of gas molecule transport through erythrocytes' membranes by kinks-solitons. *Nanosystems: Physics, Chemistry, Mathematics*. 2021;12:22-31
5. Мокрушников П.В., Рудяк В.Я. Модель диффузии липидов в цитоплазматических мембранах. *Биофизика*. 2023;68(1):41-56. doi 10.31857/S0006302923010052
   [Mokrushnikov P.V., Rudyak V.Ya. Model of lipid diffusion in cytoplasmic membranes. *Biofisica = Biophysics.* 2023;68(1):31-43. doi 10.1134/S0006350923010128 (in Russian)]
6. Yang H., Yang S., Kong J., Dong A., Yu S. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat Protoc*. 2015;10(3):382-396. doi 10.1038/nprot.2015.024

doi 10.18699/bgrs2024-3.1-18

Сайт связывания NADPH NADPH-оксидазы 2 является мишенью ингибитора GSK2795039 и его аналогов

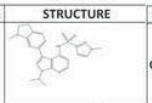
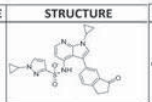
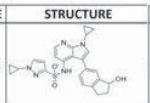
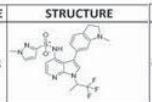
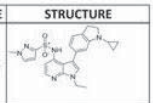
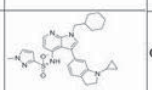
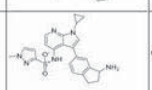
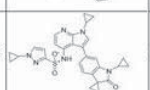
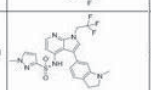
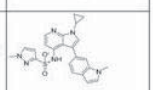
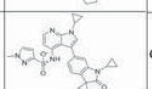
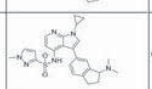
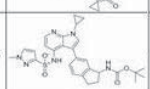
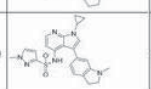
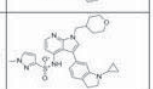
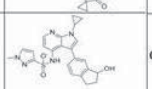
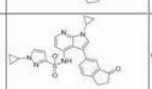
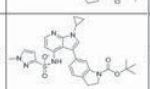
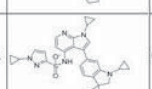
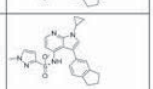
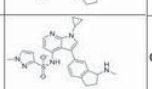
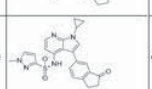
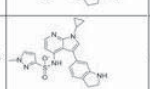
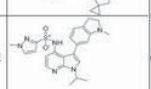
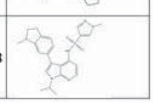
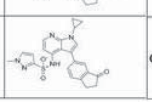
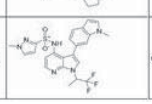
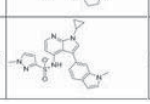
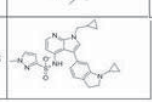
Мухина К.*, Анашкина А., Кечко О., Митькевич В.
Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия
** k.a.mukhina@mail.ru*

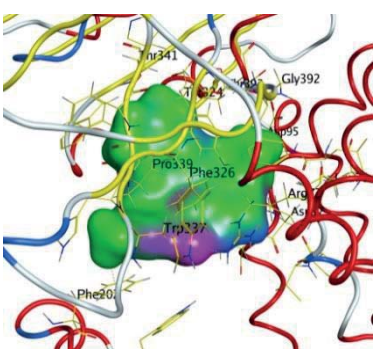
Ключевые слова: NOX2; Болезнь Альцгеймера; микроглия; GSK2795039; молекулярное моделирование; докинг; молекулярная динамика

Мотивация и цель: Болезнь Альцгеймера (БА) – наиболее распространенная нейродегенеративная патология. Недавние исследования показывают, что активация микроглии, вызванная бета-амилоидом (Aβ), играет важную роль в развитии окислительного стресса при патогенезе БА. Основным источником активных форм кислорода (АФК) в мозге является NADPH-оксидаза 2 (NOX2), которая расположена на мембране микроглиальных клеток. Мы предполагаем, что ее ингибирование, вследствие снижения окислительного стресса, может стать эффективной терапией БА. Существует селективный ингибитор NOX2 – GSK2795039 (GSK) и его аналоги [1, 2]. Целью данной работы было определение сайта связывания этих ингибиторов и оценка аффинности связывания и селективности по отношению к семейству NAD(P)H оксидаз, а также их способности ингибировать окислительный стресс в клетках микроглии человека.
Методы и алгоритмы: Структуры ингибиторов NOX2 взяты из статьи [2] (табл. 1). В качестве контроля взята структура NADPH. Исходные структуры следующих изоформ NADPH-оксидаз взяты из PDB: NOX2 (8wej), NOX5 (5o0x), DUOX1 (6wxv). Структуры оксидаз NOX1, NOX3, NOX4, DUOX2 были смоделированы программой AlphaFold (показатель достоверности (pLDDT) для NADPH связывающего участка > 95) и верифицированы методом молекулярной динамики. Молекулярная динамика проводилась программой GROMACS в течение 10 нс для более точного определения координат атомов модели. Молекулярное моделирование комплексов NOX2 и других изоформ с ингибиторами и докинг выполняли с помощью программы MOE. Для каждого лиганда с каждой из изоформ было получено по 10 конформаций взаимодействия с соответствующими значениями аффинности. Далее сравнили аффинность связывания ингибиторов в каждой изоформе и определили аффинность связывания ингибиторов в разных изоформах, т. е. селективность. Клетки микроглии человека HMC3 инкубировали 24 ч с добавлением ингибиторов GSK, GSK18, GSK28 в концентрации 25 мкМ и $A\beta_{1-42}$ 10 мкМ (Peptide Specialty Laboratories GmbH, Гейдельберг, Германия). Ингибирование окислительного стресса, вызванного Aβ, оценивали по проточной цитометрии, используя краситель на АФК дигидрородамин-123 (Ex/Em = 507/525 нм, ThermoFisher Scientific, Уолтем, Массачусетс, США). Количество мертвых клеток в популяции оценивали с помощью красителя иодида пропидия (Ex/Em = 535/617 нм; Sigma, Сент-Луис, Миссури, США).
Результаты: Были построены модельные структуры комплексов NAD(P)H оксидаз с ингибиторами. Мы определили, что исследованные ингибиторы с наибольшей аффинностью связываются с NOX2 (табл. 2). Сайт связывания

ингибиторов совпадает с сайтом связывания NADPH и включает аминокислоты Arg73, Leu76, Asp95, Asn97, Leu98, Phe202, Tyr324, Phe326, Trp337, Pro339, Thr341, Gly392, Thr393, Asn569, Phe570 (рис. 1). Были обнаружены наиболее аффинные ингибиторы: GSK1, 14, 15, 26, однако они показали наименьшую селективность (см. табл. 2). При этом за увеличение аффинности отвечают взаимодействия с аминокислотными остатками Asp95, Pro339, Phe340, Asn569, Phe570. Наилучшее действие по снижению окислительного стресса клеток показал ингибитор GSK18, достоверно снизив количество АФК на 150 % относительно клеток с Аβ и на 60 % относительно контроля (рис. 2).

Таблица 1. Структуры ингибитора NADPH-оксидазы 2 (NOX2) GSK2795039 (GSK) и его аналогов из работы [2]

NAME	STRUCTURE	NAME	STRUCTURE	NAME	STRUCTURE	NAME	STRUCTURE	NAME	STRUCTURE
GSK		GSK6		GSK12		GSK18		GSK24	
GSK1		GSK7		GSK13		GSK19		GSK25	
GSK2		GSK8		GSK14		GSK20		GSK26	
GSK3		GSK9		GSK15		GSK21		GSK27	
GSK4		GSK10		GSK16		GSK22		GSK28	
GSK5		GSK11		GSK17		GSK23			



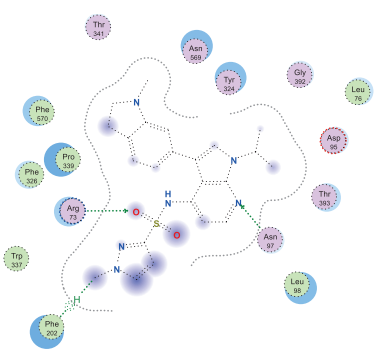


Рис. 1. Сайт связывания ингибиторов NADPH-оксидазы 2 включает аминокислоты Arg73, Leu76, Asp95, Asn97, Leu98, Phe202, Tyr324, Phe326, Trp337, Pro339, Thr341, Gly392, Thr393, Asn569, Phe570

Таблица 2. Аффинность связывания ингибиторов с NAD(P)H-оксидазами

	NOX2	NOX4	DUOX1	NOX1	NOX3	DUOX2
GSK	-8,6	-7,7	-7,1	-6,6	-6,4	-5,5
GSK1	-9,2	-7,6	-8,2	-7,1	-6,6	-6,9
GSK2	-9,0	-7,6	-7,8	-6,5	-7,1	-6,2
GSK3	-8,3	-6,7	-7,4	-6,8	-6,9	-5,9
GSK4	-9,0	-7,5	-7,5	-6,8	-7,4	-5,7
GSK5	-8,5	-7,8	-7,2	-6,3	-6,8	-5,5
GSK6	-9,0	-7,9	-7,5	-6,2	-7,0	-6,1
GSK7	-8,4	-7,3	-6,9	-6,4	-6,3	-6,4
GSK8	-8,8	-7,7	-7,5	-6,6	-6,2	-5,9
GSK9	-9,0	-7,9	-7,5	-6,2	-7,0	-6,1
GSK10	-8,5	-7,8	-7,2	-6,3	-6,8	-5,6
GSK11	-8,6	-7,6	-7,6	-6,7	-6,7	-5,4
GSK12	-8,8	-7,2	-7,5	0,0	-7,0	-6,4
GSK13	-8,8	0,0	-8,2	-6,8	-6,7	-6,4
GSK14	-9,5	0,0	-8,5	-6,1	-6,3	-5,9
GSK15	-9,2	-8,0	-7,9	-7,0	-7,0	-6,3
GSK16	-8,5	-6,8	-7,1	-6,4	-6,1	-6,4
GSK17	-8,7	-7,8	-7,3	-6,5	-6,1	-5,9
GSK18	-8,3	-6,8	-7,4	-6,3	-6,2	-5,7
GSK19	-8,5	-6,7	-7,0	-6,3	-6,8	-5,5
GSK20	-8,6	-7,3	-7,5	-6,2	-6,7	-5,6
GSK21	-9,0	-8,1	-8,2	-7,0	0,0	-6,4
GSK22	-8,6	-7,5	-7,7	-5,4	-6,6	-5,9
GSK23	-8,5	-8,5	-7,6	-6,6	-6,3	-6,8
GSK24	-8,4	-7,7	-7,4	-6,8	-6,9	-7,1
GSK25	-8,3	-7,2	-7,1	-5,9	-7,0	-6,6
GSK26	-9,2	0,0	-8,1	-6,7	-6,4	-7,1
GSK27	-8,7	-7,6	-7,3	-6,2	-6,1	-5,7
GSK28	-8,3	-7,2	-7,4	-6,3	-6,3	-5,6
NADPH	-11,6	-9,0	-9,5	-7,5	-8,0	-7,8

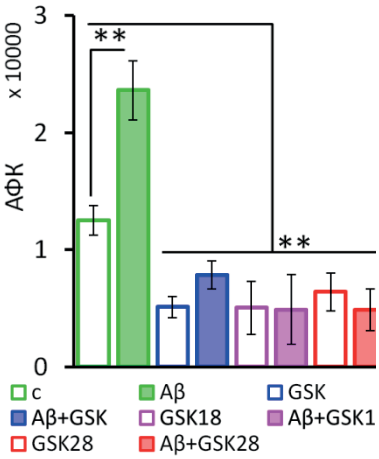


Рис. 2. Влияние ингибиторов NOX2 на уровень АФК в контрольных и Aβ-активированных клетках HMC3. Клетки обрабатывали 25 μM GSK, GSK18, GSK28 и 10 μM Aβ в течение 24 ч. Значения представляют собой среднее четырёх независимых измерений±SD. (** p<0,01)

Выводы: Впервые определены аминокислотные остатки, участвующие во взаимодействии с ингибитором GSK2795039 и его аналогами, локализованные в сайте связывания NADPH. Рассчитана аффинность связывания библиотеки ингибиторов с изоформами NAD(P)H оксидаз. Наилучший биологический эффект показал ингибитор GSK18. Установлено, что способность ингибиторов препятствовать окислительному стрессу в большей мере определяется селективностью, чем аффинностью.

Финансирование: Исследование поддержано грантом РНФ (№ 19-74-30007).

NADPH binding site NADPH oxidase 2 is a target of the inhibitor GSK2795039 and its analogues

Mukhina K.*, Anashkina A., Kechko O., Mitkevich V.

Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

** k.a.mukhina@mail.ru*

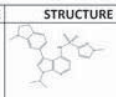
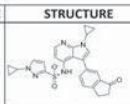
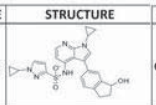
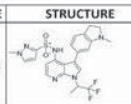
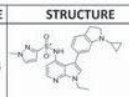
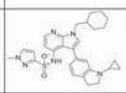
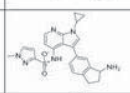
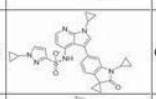
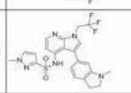
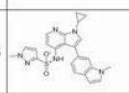
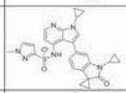
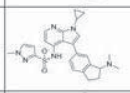
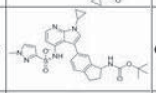
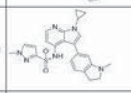
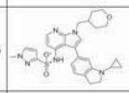
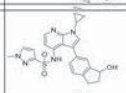
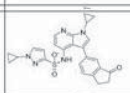
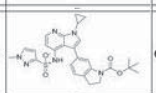
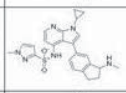
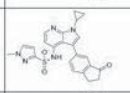
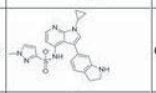
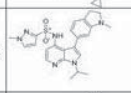
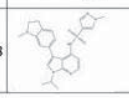
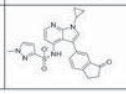
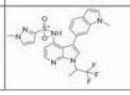
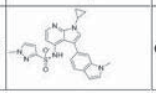
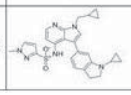
Key words: NOX2; Alzheimer's disease; microglia; GSK2795039; molecular modeling; docking; molecular dynamics

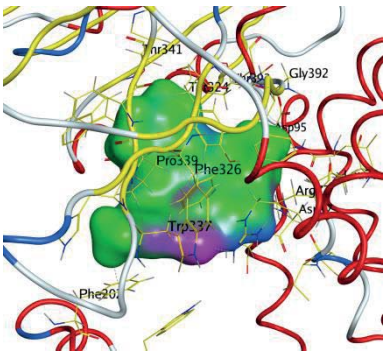
Motivation and Aim: Alzheimer's disease (AD) is the most common neurodegenerative disease. Recent studies suggest that amyloid beta (Aβ)-induced microglial activation plays an important role in the development of oxidative stress in AD pathogenesis. The major source of reactive oxygen species (ROS) in the brain is NADPH oxidase 2 (NOX2), which is located on the membrane of microglial cells. We hypothesise that its inhibition, by reducing oxidative stress, may become an effective therapy for Alzheimer's disease. There is a selective inhibitor of NOX2 – GSK2795039 (GSK) and its analogues [1, 2]. The aim of this work was to determine the binding site of these inhibitors and to evaluate their binding affinity and selectivity towards the NAD(P)H oxidase family, as well as their ability to inhibit oxidative stress in human microglial cells.

Methods and Algorithms: The structures of NOX2 inhibitors were taken from the article [2] (Table 1). The structure of NADPH was used as a control. The original structures of the following isoforms of NADPH oxidases were taken from the PDB: NOX2 (8wej), NOX5 (5o0x), DUOX1 (6wxv). The structures of the oxidases NOX1, NOX3, NOX4, DUOX2 were modelled with the program AlphaFold (confidence score (pLDDT) for the NADPH-binding site > 95) and verified by molecular dynamics. Molecular dynamics was performed using the GROMACS program for 10 nanoseconds to more accurately determine the coordinates of the model atoms. Molecular modelling of complexes of NOX2 and other isoforms with inhibitors and docking were performed using the MOE program. For each ligand with each isoform, 10 interaction conformations with corresponding affinity values were obtained. The binding affinity of the inhibitors in each isoform was then compared and the binding affinity of the inhibitors in different isoforms, i.e. selectivity, was determined. Human microglial cells HMC3 were incubated for 24 hours with the addition of GSK, GSK18, GSK28 inhibitors at a concentration of 10 μM. Inhibition of oxidative stress induced by Aβ1-42 10 μM (Peptide Specialty Laboratories GmbH, Heidelberg, Germany) was assessed by flow cytometry using the ROS dye dihydrorhodamine-123 (Ex/Em = 507/525 nm, ThermoFisher Scientific, Waltham, MA, USA). The number of dead cells in the population was assessed using propidium iodide dye (Ex/Em = 535/617 nm; Sigma, St. Louis, MO, USA).

Results: Model structures of complexes of NAD(P)H oxidases with inhibitors were constructed. We found that the inhibitors tested bind to NOX2 with the highest affinity (Table 2). The inhibitor binding site coincides with the NADPH binding site and includes the amino acids Arg73, Leu76, Asp95, Asn97, Leu98, Phe202, Tyr324, Phe326, Trp337, Pro339, Thr341, Gly392, Thr393, Asn569, Phe570 (Fig. 1). The highest affinity inhibitors were found to be: GSK1, 14, 15, 26, but they showed the lowest selectivity (Table 2). In this case, interactions with amino acid residues Asp95, Pro339, Phe340, Asn569, Phe570 are responsible for the increase in affinity. The GSK18 inhibitor showed the best effect in reducing oxidative stress in the cells, significantly reducing the amount of ROS by 150 % compared to cells with Aβ and by 60 % compared to the control (Fig. 2).

Table 1. Structures of the NADPH oxidase 2 (NOX2) inhibitor GSK2795039 (GSK) and its analogues from [2]

NAME	STRUCTURE	NAME	STRUCTURE	NAME	STRUCTURE	NAME	STRUCTURE	NAME	STRUCTURE
GSK		GSK6		GSK12		GSK18		GSK24	
GSK1		GSK7		GSK13		GSK19		GSK25	
GSK2		GSK8		GSK14		GSK20		GSK26	
GSK3		GSK9		GSK15		GSK21		GSK27	
GSK4		GSK10		GSK16		GSK22		GSK28	
GSK5		GSK11		GSK17		GSK23			



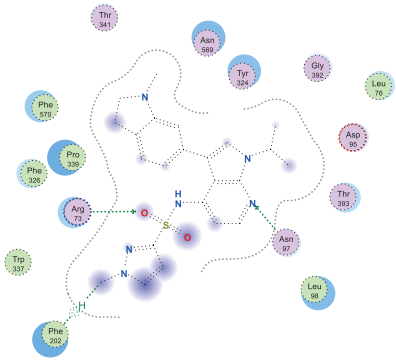


Fig. 1. The binding site for NADPH oxidase 2 inhibitors includes the amino acids Arg73, Leu76, Asp95, Asn97, Leu98, Phe202, Tyr324, Phe326, Trp337, Pro339, Thr341, Gly392, Thr393, Asn569, Phe570

Table 2. Binding affinity of inhibitors with NAD(P)H oxidases

	NOX2	NOX4	DUOX1	NOX1	NOX3	DUOX2
GSK	-8,6	-7,7	-7,1	-6,6	-6,4	-5,5
GSK1	-9,2	-7,6	-8,2	-7,1	-6,6	-6,9
GSK2	-9,0	-7,6	-7,8	-6,5	-7,1	-6,2
GSK3	-8,3	-6,7	-7,4	-6,8	-6,9	-5,9
GSK4	-9,0	-7,5	-7,5	-6,8	-7,4	-5,7
GSK5	-8,5	-7,8	-7,2	-6,3	-6,8	-5,5
GSK6	-9,0	-7,9	-7,5	-6,2	-7,0	-6,1
GSK7	-8,4	-7,3	-6,9	-6,4	-6,3	-6,4
GSK8	-8,8	-7,7	-7,5	-6,6	-6,2	-5,9
GSK9	-9,0	-7,9	-7,5	-6,2	-7,0	-6,1
GSK10	-8,5	-7,8	-7,2	-6,3	-6,8	-5,6
GSK11	-8,6	-7,6	-7,6	-6,7	-6,7	-5,4
GSK12	-8,8	-7,2	-7,5	0,0	-7,0	-6,4
GSK13	-8,8	0,0	-8,2	-6,8	-6,7	-6,4
GSK14	-9,5	0,0	-8,5	-6,1	-6,3	-5,9
GSK15	-9,2	-8,0	-7,9	-7,0	-7,0	-6,3
GSK16	-8,5	-6,8	-7,1	-6,4	-6,1	-6,4
GSK17	-8,7	-7,8	-7,3	-6,5	-6,1	-5,9
GSK18	-8,3	-6,8	-7,4	-6,3	-6,2	-5,7
GSK19	-8,5	-6,7	-7,0	-6,3	-6,8	-5,5
GSK20	-8,6	-7,3	-7,5	-6,2	-6,7	-5,6
GSK21	-9,0	-8,1	-8,2	-7,0	0,0	-6,4
GSK22	-8,6	-7,5	-7,7	-5,4	-6,6	-5,9
GSK23	-8,5	-8,5	-7,6	-6,6	-6,3	-6,8
GSK24	-8,4	-7,7	-7,4	-6,8	-6,9	-7,1
GSK25	-8,3	-7,2	-7,1	-5,9	-7,0	-6,6
GSK26	-9,2	0,0	-8,1	-6,7	-6,4	-7,1
GSK27	-8,7	-7,6	-7,3	-6,2	-6,1	-5,7
GSK28	-8,3	-7,2	-7,4	-6,3	-6,3	-5,6
NADPH	-11,6	-9,0	-9,5	-7,5	-8,0	-7,8

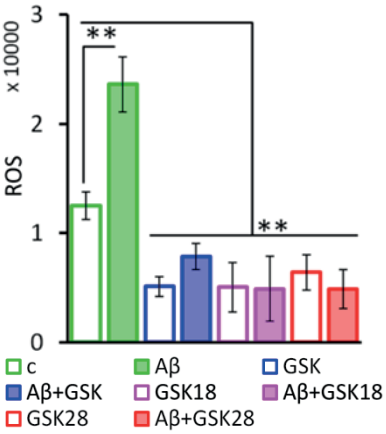


Fig. 2. Effect of NOX2 inhibitors on ROS levels in control and Aβ-activated HMC3 cells. Cells were treated with 25 μM GSK, GSK18, GSK28 and 10 μM Aβ for 24 h. Values are the mean of four independent measurements±SD. ** $p<0.01$

Conclusion: For the first time, the amino acid residues involved in the interaction with the inhibitor GSK2795039 and its analogues located in the NADPH binding site have been identified. The binding affinity of the inhibitor library to NAD(P)H oxidase isoforms was calculated. The GSK18 inhibitor showed the best biological activity. It was found that the ability of inhibitors to prevent oxidative stress is determined by selectivity rather than affinity.

Funding: The study is supported by the Russian Science Foundation (Grant No. 19-74-30007).

Список литературы/References

1. Hirano K., Chen W.S., Chueng A.L. et al. Discovery of GSK2795039, a Novel Small Molecule NADPH Oxidase 2 Inhibitor. *Antioxid Redox Signal*. 2015;23(5):358-374. doi 10.1089/ars.2014.6202
2. Mason H., Rai G., Kozyr A. et al. Development of an improved and specific inhibitor of NADPH oxidase 2 to treat traumatic brain injury. *Redox Biol*. 2023;60:102611. doi 10.1016/j.redox.2023.102611

doi 10.18699/bgrs2024-3.1-19

Функциональная аннотация генома нового вируса Хасеки (Haseki tick virus)

Осинкина И.[1], Радченко Н.[1, 2], Гладышева А.[1]*

[1] *Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора, р.п. Кольцово, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *gladysheva_av@vector.nsc.ru*

Ключевые слова: РНК-содержащие вирусы; *Flaviviridae*; вирусные белки

Мотивация и цель: Согласно данным Ежегодного отчета ВОЗ за 2023 г., количество заболеваний, вызванных инфекционными агентами, выросло почти на 2 % по сравнению с 2022 г., даже несмотря на окончание пандемии COVID-19 [1]. Изучение новых вирусов – одна из важнейших задач противодействия вирусным угрозам и обеспечения национальной биологической безопасности [2]. Однако в решении данной задачи не всегда возможно применение классических вирусологических и молекулярно-биологических подходов, и на помощь приходят методы математического моделирования, синтетической и структурной биологии. Вирус Хасеки (Haseki tick virus, HSTV) – новый вирус, впервые выявленный методом NGS в сыворотках крови людей после укуса клеща в 2020 г. на территории РФ [3]. Инфекция, вызванная HSTV, у человека характеризуется 3–5-дневной лихорадкой, затрудненным дыханием и болью в горле. Геном HSTV представлен оцРНК(+) длиной ~16 тыс. нуклеотидов и содержит одну открытую рамку считывания, фланкированную нетранслируемыми регионами. Вирусные белки синтезируются как полипротеин, посттрансляционно расщепляемый протеазами. Однако информация о том, какие вирусные белки кодирует геном HSTV и какова их пространственная организация, полностью отсутствует, что делает фактически невозможным дальнейшее предсказание и изучение биологических свойств данного вируса. Поэтому целью нашей работы стало: 1) функциональная аннотация генома вируса Хасеки; 2) получение *de novo* предсказанных вирусных белков для последующего рентгеноструктурного анализа и изучения структуры.

Методы и алгоритмы: Белковые домены HSTV были предсказаны с использованием инструмента поиска консервативных доменов NCBI, программы MOTIF в базе данных PFAM, программы InterProScan и веб-сервиса HMMER. Программа TMHMM v2.0c и веб-сервер PredictProtein.org использовались для локализации трансмембранных доменов полипротеина HSTV. Предполагаемые сайты протеолиза определяли с использованием SignalP и PROSPER. Вторичную структуру 5′–3′ UTR РНК HSTV прогнозировали с использованием серверов MFOLD 3.4 и Vienna RNA Package. Моделирование проводили при температуре 37 °C и в ионных условиях 1 M NaCl без двухвалентных ионов. Параметры «максимальное расстояние между парными основаниями» (MDBPB), «процент субоптимальности» (%S), «верхняя граница количества вычисленных складок» и «верхняя граница общего количества одноцепочечных оснований» были установлены по умолчанию, а «начальная свободная энергия ΔG» – на поиск

минимального значения. Моделирование пространственных структур белков HSTV производили с использованием AlphaFold 3 на сервере AlphaFold. Для анализа белковых структур использовали программу CrystalExplorer.

Сборку ДНК копий генов, кодирующих аминокислотные последовательности белков HSTV, осуществляли методом сборочной ПЦР с использованием высокоточной ДНК полимеразы Q5 и олигонуклеотидов длиной ~ 60 оснований и последующим клонированием продуктов амплификации в плазмидный вектор pJET. Для создания экспрессирующих генетических конструкций использовали плазмидные вектора pET200 и pEASY. Экспрессию осуществляли в штаммах BL-21(DE3) и KRX клеток *E. coli*. Растворимую фракцию рекомбинантных белков HSTV получали путем гомогенизирования клеточной биомассы ультразвуком. Клеточные лизаты анализировали путем электрофоретического разделения белков в 15 % полиакриламидном геле в денатурирующих условия по методу Лэммли.

Результаты: Было предсказано, что геном HSTV содержит одну открытую рамку считывания (ORF), кодирующую полипротеин размером ~5100 аминокислот, который ко- и посттрансляционно расщепляется на структурные и неструктурные белки (рис. 1). ORF окружена короткой 5 ′ UTR (~215 п.н.) и протяженной 3 ′ UTR (~698 п.н.). Структура полипротеина HSTV схожа с полипротеином пестивирусов. Первым белком, кодируемым HSTV, является N-концевая протеаза (Npro) размером ~168 аминокислот. Npro предположительно является цистеиновой авто-протеазой, образующей N конец капсидного белка (C) HSTV. За белком C следуют два поверхностных белка E1 и E2, содержащие трансмембранные участки.

Было предсказано, что неструктурная часть генома HSTV содержит важные NS2-подобные (протеаза), NS3-подобные (протеаза и хеликаза) и RdRp-подобные (РНК-зависимая РНК-полимераза) домены. Размер NS2-подобной протеазы около 300 а.о. Это первый белок, закодированный в неструктурной части ORF и предположительно состоящий из двух структурных доменов. Предсказанные NS3-подобные и RdRp-подобные белки составляют примерно 570 и 486 а.о. в длину соответственно. Белок NS3 состоит из домена сериновой протеазы (~210 а.о.) и консервативного C-концевого домена геликазы (~260 а.о.). Множественные трансмембранные регионы расположены между структурными белками и в конце полипротеина. Предсказанная вторичная структура 5' UTR РНК HSTV состоит из большой стеблевой петли (SLA), которая включает боковую петлю, и Y-структуры, оканчивающейся инициатором трансляции AUG. Структура SLA высококонсервативна среди различных флавивирусов, и этой структуре предшествует 5′-m7GpppAmG кэп.

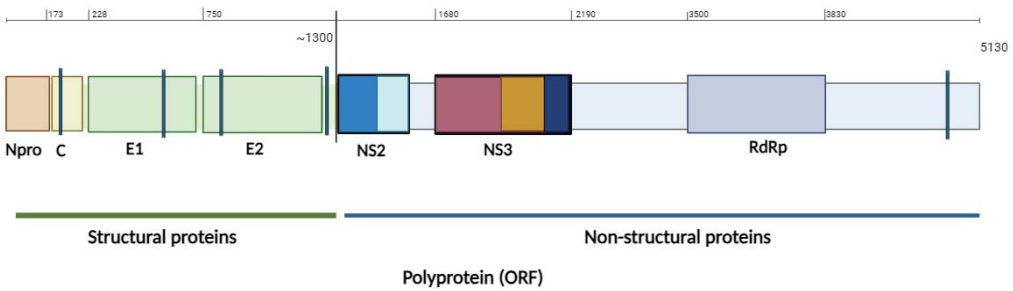


Рис. 1. Предполагаемая структура полипротеина вируса Хасеки

Модели пространственных структур белков HSTV в AlphaFold 3 имели низкий коэффициент достоверности (pLDDT Npro HSTV = 92; pLDDT SP2 HSTV = 47; pLDDT RdRp HSTV = 41) и, соответственно, не пригодны для дальнейшего анализа. Поэтому далее были получены синтетические ДНК копии генов, кодирующих предполагаемые белки Npro HSTV, надмембранную часть E2 (SP2) HSTV и RdRp HSTV, и экспрессирующие конструкции для наработки данных белков. Дизайн белков производили таким образом, чтобы получить максимальный выход белка в растворимой форме и увеличить вероятность их успешной кристаллизации. Для рекомбинантного белка Npro HSTV были созданы прокариотические штаммы-продуценты: BL-21/pET200/HSTV-Npro с оптимальными условиями наработки при культивировании в течение 16–20 часов, 20 °C и 180 об/мин с добавлением 1 мМ ИПТГ и KRX/pET200/HSTV-SP1 – культивирование 20 часов при 16 °C и 210 об/мин с индукцией 1 мМ ИПТГ и 0.1 % рамнозы.

Выводы: В данной работе мы впервые описали структуру генома нового вируса Хасеки, которая свидетельствует об эволюционном сходстве вируса Хасеки с вирусами рода *Pestiviruses*. Далее планируется кристаллизация и рентгеноструктурный анализ собранных *de novo* рекомбинантных белков Npro, SP2 и RdRp вируса Хасеки. Это позволит экспериментально подтвердить структуру белков Npro, SP2 и RdRp, установить их роль в реализации генетической информации HSTV и поможет разработать специальные тесты для обнаружения, профилактики и контроля предполагаемых инфекций HSTV у клещей и людей.

Финансирование: Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (соглашение от 12.10.2021 № 075-15-2021-1355) в рамках реализации отдельных мероприятий Федеральной научно-технической программы развития синхротронных и нейтронных исследований и исследовательской инфраструктуры на 2019–2027 гг.

Functional annotation of the novel Haseki tick virus genome

Osinkina I.[1], Radchenko N.[1, 2], Gladysheva A.[1]*

[1] *State Research Center of Virology and Biotechnology "Vector" Rospotrebnadzor, Koltsovo, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *gladysheva_av@vector.nsc.ru*

Key words: RNA viruses; *Flaviviridae*; viral proteins

Motivation and Aim: According to the WHO Annual Report for 2023, the number of diseases caused by infectious agents increased by almost 2 % compared to 2022, even despite the end of the COVID-19 pandemic [1]. The study of new viruses is one of the most important tasks in countering viral threats and ensuring national biological safety [2]. However, in solving this problem it is not always possible to use classical virological and molecular biological approaches, and methods of mathematical modeling, synthetic and structural biology come to the rescue. Haseki tick virus (HSTV) is a new virus, first identified by NGS in human blood sera after a tick bite in 2020 in the Russian Federation [3]. HSTV infection in humans is characterized by 3–5 days of fever, difficulty breathing and sore throat. The HSTV genome is represented by ssRNA(+) with a length of

~16 kilobase and contains one open reading frame flanked by untranslated regions. Viral proteins are synthesized as a polyprotein that is posttranslationally cleaved by proteases. However, information about which viral proteins the HSTV genome encodes and what their spatial organization is completely missing. It is making further prediction and study of the HSTV biological properties is impossible. Therefore, the purpose of this work was: (1) functional annotation of the Haseki virus genome and (2) obtaining *de novo* predicted viral proteins for subsequent X-ray diffraction analysis and structure studies.

Methods and Algorithms: The protein domains of HSTV were predicted using the NCBI conserved domain search tool, MOTIF Search in the PFAM database, InterProScan tool and HMMER web service. TMHMM v2.0c tool and PredictProtein.org web server were used to localize transmembrane domains for polyproteins. Putative proteolysis sites were determined using SignalP and PROSPER. The secondary structure of the 5′ UTR RNA HSTV was predicted using the MFOLD 3.4 server and Vienna RNA Package. Simulation was performed at a folding temperature of 37°C and under ionic conditions of 1 M NaCl without divalent ions. The parameters of maximal distance between paired bases (MDBPB), percent suboptimality (%S), the upper bound on the number of computed folds, and the upper bound on the total number of single-stranded bases that are allowed in a bulge or interior loop were set as default, and the initial free energy ΔG was set as the minimum value. The structures of HSTV proteins were modeled using AlphaFold 3 on the AlphaFold server. CrystalExplorer program was used to analyze protein structures.

Assembly of DNA copies of genes encoding HSTV proteins was carried out by overlap-PCR using high-precision Q5 DNA polymerase and oligonucleotides ~60 bases long, and subsequent cloning into the pJET plasmid. pET200 and pEASY vectors were used to create expression genetic constructions. Expression was carried out in strains BL-21(DE3) and KRX of *E. coli* cells. The soluble fraction of recombinant HSTV proteins was obtained by homogenizing cell biomass with ultrasound. Cell lysates were analyzed by electrophoretic separation of proteins in a 15 % polyacrylamide gel under denaturing conditions using the Laemmli method.

Results: The HSTV genome contains one open reading frame (ORF) encoding a polyprotein of ~5100 amino acids that is co- and post-translationally cleaved into structural and nonstructural proteins (Fig. 1). The ORF is flanked by a short 5' UTR (~215 bp) and a long 3' UTR (~698 bp). The HSTV polyprotein structure is similar to pestiviruses. The first protein encoded by HSTV is an N-terminal protease (Npro) of ~168 amino acids in length. HSTV Npro is a cysteine autoprotease that forms the N terminus of the HSTV capsid protein (C). Protein C is followed by two surface proteins E1 and E2 containing transmembrane regions. It was predicted that the HSTV genome contains important pestivirus NS2-like (protease), NS3-like (protease and helicase), and RdRp-like (RNA-dependent RNA polymerase) domains. The size of NS2-like protease is about 300 aa. It is the first nonstructural protein encoded in the ORF. NS2-like protease consists of two structural domains. The predicted NS3-like and RdRp-like proteins are approximately 570 aa and 486 aa in length respectively. The NS3-like protein derived into the serine protease domain (~210 aa) and the helicase conserved C-terminal domain (~260 aa). Multiple transmembrane domains located at the end of the polyprotein. The predicted secondary structure of 5′ UTR RNA HTV consists of a large stem-loop (SLA) that includes a side loop, and Y-structure terminating in the translation initiator AUG. The SLA structure is highly conserved among different flaviviruses, and this structure is preceded by a 5’-m7GpppAmG cap.

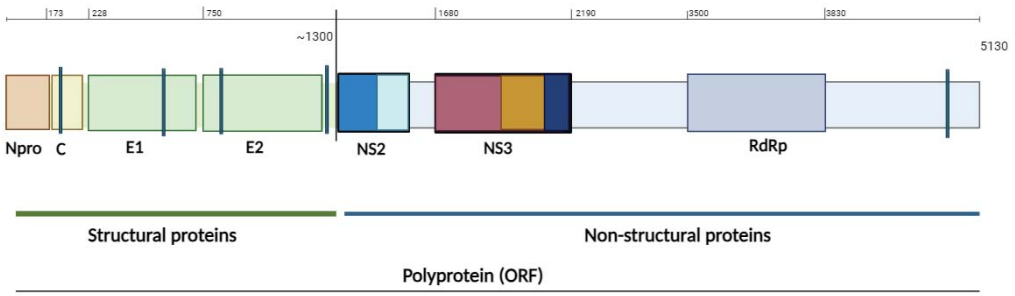


Fig. 1. Structure of the Haseki tick virus polyprotein

Models of HSTV proteins spatial structures in AlphaFold 3 had a low per-residue model confidence score (pLDDT Npro HSTV = 92; pLDDT SP2 HSTV = 47; pLDDT RdRp HSTV = 41). These protein models are not suitable for further analysis. Therefore, we obtained synthetic DNA copies of the genes encoding the putative Npro HSTV, extracellular part of E2 (SP2) HSTV and RdRp HSTV proteins, and expression constructs for the production of these proteins. Proteins were designed to maximize protein yield in soluble form and increase the likelihood of successful crystallization. Prokaryotic producer strains of recombinant Npro HSTV were created. Optimal production conditions for BL-21/pET200/HSTV-Npro are cultivation for 16-20 hours, 20 °C and 180 rpm with the addition of 1 mM IPTG and for KRX/pET200 /HSTV-SP1 are cultivation for 20 hours at 16 °C and 210 rpm with induction of 1 mM IPTG and 0.1 % rhamnose.

Conclusion: In this work, we described for the first time the genome structure of the new Haseki tick virus. This result indicates the evolutionary similarity of the Haseki tick virus with viruses of the *Pestiviruses* genus. Next, crystallization and X-ray diffraction analysis of the *de novo* assembled recombinant Npro, SP2, and RdRp Haseki tick virus proteins are planned. This will allow us to experimentally confirm the structure of the Npro, SP2 and RdRp proteins, establish their role in the implementation of HSTV genetic information and help to develop specific tests for the detection, prevention and control of putative HSTV infections in ticks and humans.

Funding: The study was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2021-1355 dated October 12, 2021) as a part of the implementation of certain activities of the Federal Scientific and Technical Program for the Development of Synchrotron and Neutron Research and Research Infrastructure for 2019–2027.

Список литературы/References

1. https://www.who.int/data/gho/publications/world-health-statistics
2. Ergunay K. et al. The expanding range of emerging tick-borne viruses in Eastern Europe and the Black Sea Region. *Sci Rep*. 2023;13(1):19824. doi 10.1038/s41598-023-46879-2
3. Kartashov M.Y., Gladysheva A.V., Shvalov A.N., Tupota N.L., Chernikova A.A., Ternovoi V.A., Loktev V.B. Novel Flavi-like virus in ixodid ticks and patients in Russia. *Ticks Tick Borne Dis.* 2023;14(2):102101. doi 10.1016/j.ttbdis.2022.102101

doi 10.18699/bgrs2024-3.1-20

Структурные основы функциональных дефектов митохондриальной пирофосфатазы PPA2 при мутациях, вызывающих кардиопатологии человека

Родина Е.В.[1]*, Безпалая Е.Ю.[1], Курилова С.А.[1, 2], Воробьева Н.Н.[2]

[1] *Химический факультет, Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

[2] *НИИ ФХБ имени А.Н. Белозерского, МГУ им. М.В. Ломоносова, Москва, Россия*

* *rodina@belozersky.msu.ru*

Ключевые слова: неорганическая пирофосфатаза; митохондрия; структура; мутация; конформационные перестройки

Мотивация и цель: Неорганические пирофосфатазы (PPазы) катализируют гидролиз пирофосфата, обеспечивая протекание важнейших биосинтетических реакций. Ядерный геном животных и дрожжей кодирует две различные формы пирофосфатазы: PPA1 (цитоплазматическая) и PPA2 (митохондриальная). Цитоплазматические PPазы и их гомологи в микроорганизмах – хорошо изученные ферменты, в то время как о строении, функции и биологической роли митохондриальных гомологов известно очень мало. Из данных литературы известно, что PPA2 дрожжей *S. cerevisiae* необходима для аэробного метаболизма и сохранения целостности митохондриальной ДНК [1]. PPA2 человека, по-видимому, также играет критически важную роль в функционировании митохондрий. Так, природные биаллельные мутации в гене *ppa2* приводят к митохондриальным кардиопатологиям, для которых характерны дефекты в экспрессии и/или каталитической активности PPA2 [2, 3]. При этом мутированные остатки не имеют очевидной связи с функционально важными элементами белка, а спектр и тяжесть клинических проявлений варьируют у различных пациентов без корреляции с конкретными мутациями. Для выяснения молекулярных основ данного заболевания необходимо более полное понимание метаболической роли PPA2, ее регуляции в клетке, а также влияния конкретных мутаций на структуру и функцию фермента. Целью настоящей работы является характеристика *in vitro* каталитических свойств PPA2, получение и анализ структуры PPA2, а также характеристика изменений структуры и функции PPA2 при мутациях, соответствующих патогенным заменам фермента человека.

Методы и алгоритмы: Для функциональной характеристики PPA2 и мутантных вариантов *in vitro* исследовали стационарную кинетику гидролиза $MgPP_i$. Связывание лигандов и оценку влияния мутаций на структуру рекомбинантных белков характеризовали стандартным набором биохимических и расчетных методов, включая термофлуориметрию, изотермическую калориметрию и молекулярный докинг. Структурная характеристика PPA2 включала анализ полученной в рамках работы кристаллической структуры PPA2 из *O. parapolymorpha* (OpPPA2) и модели структуры PPA2 человека (hPPA2). Поведение структуры в растворе и влияние на нее мутаций исследовали с помощью симуляции молекулярной динамики.

Результаты: В работе были определены функциональные параметры hPPA2: каталитические параметры V_{max} и K_m, константы диссоциации комплексов ферментов с кофакторными ионами Mg^{2+}, константы ионизации каталитически важных групп фермента и фермент-субстратного комплекса. Также в работе исследован мутантный вариант Met94Val, соответствующий патогенному природному варианту. Этот вариант вызывает интерес, поскольку трудно представить, каким образом гомологичная замена гидрофобного остатка на такой же гидрофобный остаток приводит к столь серьезной митохондриальной дисфункции. Ранее в литературе было выдвинуто предположение, что такая замена приводит к дестабилизации белковой структуры [3], однако в нашей работе показано, что данная мутация не снижает стабильность белка, а изменяет именно его каталитические свойства. Согласно нашим результатам, в рекомбинантном белке замена Met94Val значительно снижает V_{max} и приводит к увеличению pK_a каталитического основания, необходимого для активации молекулы воды при образовании атакующего нуклеофила. Также значительно снижается сродство фермента к ингибитору – фторид-иону, имитирующему атакующий нуклеофил. Аналогичная замена в OpPPA2 (Met52Val) характеризуется такими же функциональными последствиями и дополнительно приводит к существенному снижению сродства фермента к кофактору Mg^{2+} и субстрату $MgPP_i$. Таким образом, причиной функциональных дефектов при мутации Met94Val (Met52Val в OpPPA2), вероятно, является нарушение образования атакующего нуклеофила. При этом изменения глобальной структуры белка в результате данной мутации не происходит.

Ранее в работе была получена кристаллическая структура OpPPA2 в холоформе (комплексе с кофакторными ионами Mg^{2+}) с разрешением 1.8 Å. Анализ этой структуры, а также модельной структуры hPPA2 показывает, что остаток Met52 (как и аналогичный ему Met94 в hPPA2) расположен в гидрофобном коре белковой глобулы и участвует в заякоривании пептидных групп каталитически важных остатков активного центра – лигандов кофакторных Mg^{2+} и субстрата (остатки Glu97 и Tyr142 в hPPA2). Замена Met на более короткий Val может ухудшать связи между этими каталитическими остатками и гидрофобным кором, тем самым приводя к нарушению пре-организации активного центра и, как следствие, к снижению каталитической эффективности. Симуляция молекулярной динамики, выполненная как на OpPPA2 и его мутантном варианте Met52Val, так и на белке человека hPPA2 и его мутантном варианте Met94Val, подтверждает это предположение. Кроме того, обнаружено, что в результате данной мутации в обоих ферментах снижается общая подвижность полипептидной цепи. Анализ полученных в динамике данных методом PCA (анализа главных компонент) также показывает, что характер конформационных изменений изменяется в результате мутации. Согласно данным литературы [4], замена метионина на валин часто снижает структурную адаптабельность белкового кора, которая требуется для функционально важных конформационных изменений; вероятно, в случае PPA2 это приводит к затруднению тех перестроек, которые необходимы для получения и активации атакующего нуклеофила.

Анализ других остатков, мутации которых в hPPA2 приводят к кардиопатологиям, показывает, что консервативные и полуконсервативные позиции (Ser61, Met106, Arg127, Pro167, Glu172, Pro228) сохраняются и в цитоплазматических гомологах PPA1. Предположительно, их роль в структуре и функции белка не специфична

именно для митохондриальной формы фермента, а их замены имеют схожие проявления во всех гомологичных PPазах, как это наблюдается для Met94. Симуляция молекулярной динамики позволяет предсказать, какое влияние эти мутации могут оказывать на структуру и свойства белка.
Кристаллическая структура OpPPA2 и модельная структура hPPA2 также были использованы для картирования группоспецифичных аминокислотных остатков и анализа структурных детерминант, отличающих митохондриальный фермент от цитоплазматических гомологов. Согласно этому анализу, структура PPA1 более компактна, в то время как в PPA2 содержатся дополнительные структурные элементы (в основном петли), которые, по-видимому, участвуют в молекулярном узнавании PPA2 клеточными партнерами. Группоспецифичные аминокислотные остатки, отличающие PPA2 от PPA1, расположены в основном на поверхности белка и кластеризуются в области, которая предсказана как аллостерический центр, что также может иметь отношение к функциональным различиям PPA2 и PPA1.
Выводы: Проведенный анализ позволил выявить структурные основы каталитической несостоятельности мутантного варианта Met94Val в митохондриальной пирофосфатазе человека hPPA2. Полученные данные также позволяют предсказать функциональные дефекты и в результате других замен, соответствующих патогенным мутантным вариантам hPPA2.
Финансирование: Исследование поддержано грантом РНФ (№ 23-24-00177).

Structural basis for the functional defects in mitochondrial inorganic pyrophosphatase PPA2 caused by mutations responsible for human cardio pathologies

Rodina E.[1]*, Bezpalaya E.[1], Kurilova S.[1, 2], Vorobyeva N.[2]

[1] *Lomonosov Moscow State University, Department of Chemistry, Moscow, Russia*

[2] *Lomonosov Moscow State University, A.N. Belozersky Institute of Physico-Chemical Biology, Moscow, Russia*

* *rodina@belozersky.msu.ru*

Key words: inorganic pyrophosphatase; mitochondria; structure; mutation; conformational changes

Motivation and Aim: Inorganic pyrophosphatase (PPase) catalyze hydrolysis of inorganic pyrophosphate thereby supporting the key biosynthetic reactions. The nuclear genome of animals and yeast codes for the two forms of pyrophosphatase: cytoplasmic (PPA1) and mitochondrial (PPA2). Cytoplasmic pyrophosphatases and their homologs from microorganisms are well studied enzymes, whereas the knowledge on the structure, function, and metabolic role of mitochondrial homologs is scarce. PPA2 from yeast *S. cerevisiae* is necessary for the aerobic metabolism and maintaining mitochondrial DNA [1]. In human, PPA2 is thought to be critically important for the mitochondrial function. Biallelic mutations in the gene *ppa2* are responsible for the mitochondrial cardio pathologies with impaired expression and/or catalytic activity of PPA2 [2, 3]. The affected patients demonstrate a wide spectrum of clinical manifestations of varying severity without correlation with particular mutations, while mutated residues have no obvious connection to the functionally important protein elements. Identification of the

molecular basis of this disease requires a detailed understanding of PPA2 metabolic role, regulation, and the impact of particular mutations on the enzyme structure and function. This work is aimed at the *in vitro* characterization of the catalytic properties of PPA2, analysis of its structure, and the characterization of the effects of mutations responsible for human cardio pathologies on the protein structure and function.

Methods and Algorithms: *In vitro* functional characterization of PPA2 and its mutant variants included steady-state kinetics of $MgPP_i$ hydrolysis. Ligand binding and the effects of mutations on the recombinant proteins structure was studied by a standard set of biochemical and computational methods including Thermofluor, ITC, molecular docking etc. Structural characterization of PPA2 included the analysis of a crystal structure of *O. parapolymorpha* PPA2 (OpPPA2) obtained earlier in this work, and of the modelled structure of human PPA2 (hPPA2). The solution dynamics of these structures and the effects of mutations was studied by the molecular dynamic simulation.

Results: In this work we determined functional parameters of hPPA2: catalytic parameters V_{max} and K_m, dissociation constants of the enzyme complexes with cofactor Mg^{2+}, ionization constants of the catalytically important ionizable groups of the enzyme and enzyme-substrate complex. Mutant variant Met94Val corresponding to a pathogenic natural variant has also been characterized. This variant was of especial interest: this is unclear how exactly a homologous substitution of a non-polar residue with another non-polar residue can cause a severe mitochondrial disfunction. It has earlier been suggested [3] that this mutation destabilizes protein structure; however, we found in the present work that this mutation affects catalytic properties of PPase without global changes in its structure or stability. We demonstrate here that the mutation Met94Val in recombinant enzyme significantly decreased V_{max} and increased pK_a of catalytic base required for the activation of a water molecule to become an attacking nucleophile. Another effect of mutation is a drastically decreased affinity for the fluoride ion inhibiting PPases by mimicking the attacking nucleophile. Analogous mutation in OpPPA2 (Met52Val) is characterized by similar functional defects; in addition, affinity for the cofactor Mg^{2+} and substrate $MgPP_i$ is drastically decreased in this variant. Therefore, the impaired formation of the attacking nucleophile is presumably the leading cause of a functional incompetence of Met94Val (Met52Val in OpPPA2), while there is no evidence of this mutation causing global structural changes.

Within this work we have earlier obtained a crystal structure of OpPPA2 in a holoform (i. e. complexed with two cofactor Mg^{2+} ions) at 1.8 Å resolution. Analysis of this structure as well as of a model structure of hPPA2 demonstrates that Met94 in hPPA2 (similarly to Met52 in OpPPA2) is located in the protein core inside a beta-barrel and participated in anchoring peptide groups of catalytically important residues of the active site (Glu97 and Tyr142 in hPPA2) which coordinate cofactor Mg^{2+} and substrate. Substitution of Met with shorter Val may disrupt the connections between these catalytic residues and the protein core, thereby distorting the pre-organized active site architecture and as a consequence decreasing catalytic efficiency. Molecular dynamic simulation performed on four proteins (wild-type hPPA2 and Met94Val, wild-type OpPPA2 and Met52Val) confirmed this supposition. We additionally found that this mutation in both proteins caused the decrease in the global flexibility of a polypeptide chain. Analysis of molecular dynamics data by PCA (principal Component Analysis) showed that mutation changes the mode of conformational dynamics of a protein in solution. According to the modern views in protein structure, substitution of methionine with valine or other branched-chain aliphatic residues decrease the structural adaptability of protein core

required for the functionally important conformational rearrangements [4]. We suppose that in the case of PPA2 such a mutation may impair the rearrangements required for the formation and activation of an attacking nucleophile.

Analysis of other residues related to the pathogenic mutations in hPPA2 demonstrates that residues same or similar in both hPPA2 and OpPPA2 (Ser61, Met106, Arg127, Pro167, Glu172, Pro228) are also conserved in cytoplasmic homologs, PPA1. Presumably, their role in PPase structure and function is not specific for mitochondrial form PPA2, but rather their mutations would similarly affect all homologous PPases, as we observe for Met94 mutation. Molecular dynamic simulation allows an insight into these effects on the protein structure and function.

Crystal structure of OpPPA2 and a model structure of hPPA2 have also been used in this work for the mapping of group-specific amino acid residues and the analysis of the structural determinants that distinguish mitochondrial enzyme from cytoplasmic homologs. This analysis shows that PPA1-type structure is more compact while PPA2-type structures have additional structural elements (mostly loops) which, presumably, are involved in molecular recognition of PPA2 by cell partners. Most of group-specific amino acid residues distinct in PPA2-type and PPA1-type PPases are located at the protein surface and are clustered at a region predicted to be an allosteric site. This can also be related to a functional distinction between PPA2 and PPA1 proteins.

Conclusion: The combined analysis performed in the work allows an insight into the structural basis of the catalytic incompetence of a mutant variant Met94Val in human mitochondrial pyrophosphatase hPPA2. The data obtained in the work allow the prediction of functional defects caused by another pathogenic mutant variants of hPPA2.

Funding: The study is supported by the Russian Science Foundation (No. 23-24-00177).

Список литературы/References

1. Lundin M., Baltscheffsky H., Ronne H. Yeast PPA2 gene encodes a mitochondrial inorganic pyrophosphatase that is essential for mitochondrial function. *J Biol Chem*. 1991;266:12168-12172
2. Guimier A., Gordon C.T., Godard F. et al. Biallelic PPA2 mutations cause sudden unexpected cardiac arrest in infancy. *Am J Hum Genet*. 2016;99:666-673. doi 10.1016/j.ajhg.2016.06.021
3. Kennedy H., Haack T.B., Hartill V. et al. Sudden cardiac death due to deficiency of the mitochondrial inorganic pyrophosphatase PPA2. *Am J Hum Genet*. 2016;99:674-682
4. Aledo J.C. Methionine in proteins: The Cinderella of the proteinogenic amino acids. *Protein Sci*. 2019;28(10):1785-1796. doi 10.1002/pro.3698

doi 10.18699/bgrs2024-3.1-21

Лигирование OaAEP1 и SrtA в мембраноподобном окружении

Савицкая А.*, Гончарук С.
ФГБУН «Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН», Москва, Россия
** sanna3061@gmail.com*

Ключевые слова: лигирование; мембранные белки; OaAEP1; SrtA; мембраноподобное окружение

Мотивация и цель: Селективное лигирование двух или более белков-субстратов представляет подход, который позволяет получить продукт в обход рибосомального синтеза. Такой метод синтеза позволяет облегчить получение токсичных полипептидов большого размера, белков с трансмембранными доменами и сделать возможным сегментное мечение доменов полипептидов, которое позволит охарактеризовать структурные характеристики белка с помощью ЯМР. Одним из методов лигирования является использование ферментов аспарагинилэндопептидазы (OaAEP1) и сортазы А (SrtA). В данной работе мы проводим сравнение активности лигирования OaAEP1 и SrtA при лигировании водорастворимых белков и белков-субстратов, один из которых является тремнсмембранным и требует наличия мембранного окружения.
Методы и алгоритмы: Пептидазы AEP1 из гена *Oldenlandia affinis* (OaAEP1) (UniProtKB: A0A0N9JZ32_OLDAF) и кальций-независимый мутант сортазы А *Staphylococcus aureus* SrtA (Addgene: 105602) были экспрессированы в клетках BL21(DE3) *E. coli* и очищены с помощью металлохелатной хромотографии [1, 2]. Очищенный OaAEP1 был активирован перед проведением лигирования [3]. Конструкции модельных белков-субстратов Smt3 и p75TM содержат “NAL” на С-конце, p75ICD содержит “GL” на N-конце для проведения лигирования с помощью OaAEP1. Для лигирования с помощью SrtA конструкции Smt3 и p75TM содержали “LPETG” на С-конце, p75ICD содержал “GG” на N-конце. Модельные белки Smt3 и p75ICD также были получены в клетках BL21(DE3) *E. coli* и очищены с применением металлохелатной хроматографии. P75TM был получен в бесклеточной среде и очищен с помощью гельфильтрации в 0.5 % детергента лаурилсаркозинат натрия и переосажден [4]. Для проведения лигирования белки-субстраты смешивали в равном соотношении (0.02 мМ). В качестве мембраноподобного окружения были протестированы мицеллы, образованные DDM, CHAPS и бицеллы DMPC/CHAPS, DMPC/DHPC, DMPC/facade-EM, DMPC/facade-TEM, DMPC/facade-EPC, DMPC/facade-TEG, DMPC/facade-TFA1.
Результаты: Пептидазы SrtA и OaAEP1 имели схожую эффективность лигирования водорастворимых белков-субстратов Smt3 и p75ICD, которая не превышала 50 % от белков-субстратов. Однако при лигировании SrtA наблюдалось более интенсивное образование дополнительных продуктов лигирования по сравнению с OaAEP1. Лигирование трансмембранного белка-рецептора нейротрофина $p75^{NTR}$ (p75TM) с его внутриклеточным доменом (p75ICD) проводилось в мембранном окружении. Наибольший выход лигированного p75TM-ICD, соспоставимый с выходом при лигировании водорастворимых белков-субстратов, был получен в

мицеллах, сформированных DDM при лигировании OaAEP1. Наименьшее количество лигированного продукта при лигировании OaAEP1 было в мицеллах, образованных CHAPS, а также бицеллами DMPC/CHAPS, DMPC/DHPC и DMPC/фасад-EPC. Эффективность лигирования сортазой A (SrtA) в мембрано-подобной среде была ниже по сравнению с OaAEP1.
Финансирование: Исследование поддержано грантом РНФ (№ 22-14-00130).

OaAEP1 and SrtA ligation in a membrane-like environment

Savitskaya A.*, Goncharuk S.
Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia
** sanna3061@gmail.com*

Key words: ligation; membrane proteins; OaAEP1; SrtA

Motivation and Aim: Selective ligation of two or more substrate proteins represents an approach that bypasses ribosomal synthesis to obtain proteins. Ligation makes it easier to obtain large toxic polypeptides, proteins with transmembrane domains, and segmentally labeling of polypeptide domains, which will allow characterizing the structural characteristics of the protein by NMR. One of the ligation methods is the use of the enzymes as asparaginyl endopeptidase (OaAEP1) and sortase A (SrtA). We compare the efficiency of ligation by OaAEP1 and SrtA by using as an substrates a water-soluble proteins and membrane protein that requires a membrane environment.
Methods and Algorithms: AEP1 from the *Oldenlandia affinis* (OaAEP1) (UniProtKB: A0A0N9JZ32_OLDAF) and calcium-independent sortase A mutant Staphylococcus aureus SrtA (Addgene: 105602) were expressed in BL21(DE3) *E. coli* cells and purified by IMAC [1, 2]. Constructs of substrate proteins Smt3 and p75TM contains “NAL” at the C-terminus, p75ICD contains “GL” at the N-terminus for ligation by OaAEP1 [3]. Ligation by SrtA is generated with constructs Smt3 and p75TM contained “LPETG” at the C-terminus, p75ICD contained “GG” at the N-terminus. Smt3 and p75ICD were also produced in BL21(DE3) *E. coli* competent cells and purified using IMAC. P75TM was expressed by *cell-free protein synthesis* and purified by gel filtration at 0.5 % lauryl sarcosinate and precipitated [4]. The substrate proteins for ligation were mixed in equal proportions (0.02 mM). Membrane mimic environment were formed by DDM, CHAPS and bicelles DMPC/CHAPS, DMPC/DHPC, DMPC/facade-EM, DMPC/facade-TEM, DMPC/facade-EPC, DMPC/facade-TEG, DMPC/facade-TFA1.
Results: The peptidases SrtA and OaAEP1 had similar ligation efficiency in case of water-soluble substrate proteins Smt3 and p75ICD ligation, which did not exceed 50 % of the substrate proteins. However, upon ligation by SrtA, a more intense formation of additional ligation products was observed compared to OaAEP1. Ligation of the transmembrane neurotrophin receptor protein $p75^{NTR}$ (p75TM) with its intracellular domain (p75ICD) was carried out in a membrane-mimic environment. The highest yield of ligated p75TM-ICD, comparable to the yield from ligation of water-soluble substrate proteins, was obtained in micelles formed by DDM at ligation generated by OaAEP1. The smallest amount of ligated product during OaAEP1 ligation was in micelles formed by CHAPS, as well as DMPC/CHAPS, DMPC/DHPC and DMPC/facade-EPC bicelles.

The efficiency of ligation by sortase A (SrtA) in the membrane-like environment of BA is lower compared to OaAEP1.

Funding: The study is supported by Russian Science Foundation (No. 22-14-00130).

Список литературы/References

1. Harris K., Durek T., Kaas Q. et al. Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat Commun.* 2015;6:10199. doi 10.1038/ncomms10199
2. Wu Z., Guo X., Gao J., Guo Z. Sortase A-mediated chemoenzymatic synthesis of complex glycosylphosphatidylinositol-anchored protein. *Chem Commun.* 2013;49(99):11689
3. Dall E., Brandstetter H. Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proc Natl Acad Sci USA.* 2013;110(27):10940-10945. doi 10.1073/pnas.1300686110
4. Schwarz D., Junge F., Durst F. et al. Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nat Protoc*. 2007;2:2945-2957. doi 10.1038/nprot.2007.426

doi 10.18699/bgrs2024-3.1-22

Структурное исследование апо- и холо-CopC из *Thiolkalivibrio paradoxus*

Соловьева А.[1]*, Варфоломеева Л.[1], Куликова О.[1], Тихонова Т.[1], Попов В.[1, 2]

[1] *ФГУ Федеральный исследовательский центр «Основы биотехнологии» РАН, Москва, Россия*

[2] *Московский государственный университет им. Ломоносова, Москва, Россия*

** nastya.soloveva1@yandex.ru*

Ключевые слова: медь; медь-связывающий белок; CopC, металлошапероны; рентгеноструктурный анализ

Мотивация и цель: Семейство белков CopC (copper resistant protein) представляет собой периплазматические медь-связывающие белки, которые играют важную роль в гомеостазе меди. Изначально CopC рассматривались как один из компонентов системы устойчивости к ионам меди CopABCDERS. Но позже было обнаружено, что часто ген CopC присутствует в генных кластерах только с геном белка внутренней мембраны CopD, а также с генами медных оксидаз [1]. Предполагается, что CopC в данном случае выполняет функцию металло-шаперона, участвующего во встраивании ионов меди в активный центр медных ферментов. В геноме бактерии *Thiolkalivibrio paradoxus* гены CopCD входят в генный кластер, отвечающий за биосинтез медь-содержащего белка тиоцинатдегидрогеназы (TcDH) [2]. Целью настоящей работы являлось структурное исследование медь-связывающего белка CopC из *Tv. paradoxus* – потенциального металлошаперона TcDH, а также анализ структурных отличий апо- и холоформ CopC.

Методы и алгоритмы: В работе были использованы методы генной инженерии для получения рекомбинантной апо-формы CopC, хроматографические методы для выделения и очистки CopC и биохимические методы для его характеристики. Холоформу получали насыщением апоформы CopC ионами Cu^{2+} с последующим удалением избытка ионов меди диализом. Кристаллы CopC получали методом диффузии паров, структуры CopC решены методом рентгеноструктурного анализа.

Результаты: Получены и уточнены структуры апо- и холо-CopC с разрешением 1.70 и 1.80 Å. В обеих структурах CopC находится в форме димера, что отличает CopC из *Tv. paradoxus* от ранее описанных гомологов из других бактерий, которые находятся в мономерном состоянии. В образовании димера участвуют C-концы мономеров («хвост к хвосту»), на N-конце каждого мономера формируется характерный для CopC медь-связывающий мотив, который в литературе называется «гистидиновая скрепка». Также в структуре CopC из *Tv. paradoxus* имеется вставка из 28 аминокислот, которая отсутствует у гомологичных CopC [1]. Данная вставка формирует петлю и α-спираль, и ее функция, возможно, заключается в формировании сайта для взаимодействия с белками-партнерами. Это подтверждает полученная ранее структура комплекса CopC-FCC, где флавиновая субъединица флавоцитохромсульфиддегидрогеназы (FCC) взаимодействует с данной α-спиралью [3].

В структуре холоформы *Tv. paradoxus* CopC сайт, на котором связан ион Cu^{2+}, имеет строение квадратной пирамиды и формируется N-концевым гистидином и двумя консервативными остатками – Asp110 и His112. Такое же строение имеют медь-связывающие сайты CopC из других бактерий [1].
В структуре апоформы в области «гистидиновой скрепки» также присутствовал ион меди с заселенностью 0.3. Остатки Asp110 и His112 не меняют своего положения в сравнении с холоформой, сохраняя сеть образуемых водородных связей, тогда как N-концевой гистидин либо не виден, либо его аминогруппа отвернута и не принимает участие в координации меди. В ходе всей полипептидной цепи также наблюдается ряд различий между апо- и холо-CopC: петля 80–83, которая взаимодействует с участком, содержащим консервативный Glu28 из вторичной координационной сферы иона меди, имеет разные положения.
Выводы: CopC из *Tv. paradoxus* обладает рядом структурных отличий от ранее описанных гомологов: во-первых, является димером, во-вторых, имеет дополнительную вставку из 28 аминокислот, которая участвует в белок-белковых взаимодействиях. Помимо этого, были найдены и описаны различия в структуре CopC в присутствии/отсутствие ионов меди: во-первых, координирующий His1 либо не описывается электронной плотностью, либо развернут от меди, во-вторых, отличается положение петли 80–83.
Финансирование: Исследование поддержано грантом Российского научного фонда № 23-74-30004.

Structural study of apo- and holo-CopC from *Thiolkalivibrio paradoxus*

Solovieva A.[1]*, Varfolomeeva L.[1], Kulikova O.[1], Tikhonova T.[1], Popov V.[1, 2]
[1] *Research Centre of Biotechnology, RAS, Moscow, Russia*
[2] *Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia*
* *nastya.soloveva1@yandex.ru*

Key words: copper; copper-binding protein; CopC; metallochaperones; X-ray diffraction analysis

Motivation and Aim: The CopC (copper resistant proteins) family is periplasmic, copper-binding proteins that play an important role in copper homeostasis. CopC was initially considered as one of the components of the copper resistance system – CopABCDERS. But later it was found that CopC is often present in gene clusters only with the inner membrane protein CopD, and also genes of cuproenzymes [1]. It is assumed that CopC in this case functions as a metallochaperone, which binds copper ions and then incorporates in the active center of cuproenzymes. In the genome of the bacterium *Thiolkalivibrio paradoxus*, the CopCD genes are included in the gene cluster responsible for the biosynthesis of the copper-containing thiocyanate dehydrogenase (TcDH) [2]. In this work, the goal was to obtain structure of CopC – a potential metallochaperone of TcDH, and to identify structural changes in apo- and holo-forms.
Methods and Algorithms: In the work we used genetic engineering methods to obtain the recombinant apo-CopC, chromatographic purification methods for isolation and purification of CopC, and biochemical methods for its characterization. The holo-CopC

was obtained by saturating the apo-CopC with Cu^{2+}, followed by removal of excess copper ions by dialysis. CopC crystals were obtained by vapor diffusion; the CopC structures were solved by X-ray diffraction.

Results: The structures of apo- and holo-CopC were obtained and solved with resolution 1.70 and 1.80 Å. In both structures, CopC was a dimer that is in contrast to the structures of previously described monomeric homologs. The C-termini of the monomers ("tail to tail") form a dimer interface: and the N-terminus of each monomer forms a copper-binding motif, which is called the histidine brace in the literature. Also CopC from *Tv. paradoxus* has a 28 amino acid insertion that is not present in other homologous CopCs [1]. This insertion forms a loop and α-helix and acts as a platform for interaction with partners-proteins. This is confirmed by the previously obtained structure of the CopC-FCC complex, where the flavin subunit of flavocytochrome sulfide dehydrogenase (FCC) interacts with this α-helix [3].

In the structure of the holo-CopC from *Tv. paradoxus* copper-binding site has the structure of a square pyramid and is formed by the N-terminal histidine and two conservative residues – Asp110 and His112. The copper-binding sites of CopC from other bacteria have the same structure [1].

In the structure of the apo-CopC, in the histidine brace there was also a copper ion with an occupancy of 0.3. Residues Asp110 and His112 do not change their position in comparison with the holo-form, maintaining the network of formed hydrogen bonds. While the N-terminal histidine is either not visible, or its amino group is turned away and does not participate in the coordination of copper. There is also a difference in the polypeptide chain position between apo- and holo-CopC: loop 80–83, which interacts with the region containing the conserved Glu 28 from the secondary copper coordination sphere, has different locations.

Conclusion: CopC from *Tv. paradoxus* has a number of structural differences from previously described homologs: firstly, it is a dimer, and secondly, it has an additional insert of 28 amino acids, which is involved in protein-protein interactions. In addition, differences in the structure in the presence/absence of copper were found and described: firstly, the coordinating His1 is either not described by the electron density or turns away from copper, and secondly, the position of loop 80–83 is different.

Funding: The study is supported by the Russian Science Foundation grant No. 23-74-30004.

Список литературы/References

1. Lawton T.J., Kenney G.E., Hurley J.D., Rosenzweig A.C., The CopC Family: Structural and Bioinformatic Insights into a Diverse Group of Periplasmic Copper Binding Proteins. *Biochemistry*. 2016;55(15):2278-2290. doi 10.1021/acs.biochem.6b00175
2. Berben T., Overmars L., Sorokin D.Y., Muyzer G., Comparative genome analysis of three thiocyanate oxidizing Thioalkalivibrio species isolated from soda lakes. *Front Microbiol*. 2017;8:254. doi 10.3389/fmicb.2017.00254
3. Osipov E.M., Lilina A.V., Tsallagov S.I., Safonova T.N., Sorokin D.Y., Tikhonova T.V., Popov V.O. Structure of the flavocytochrome c sulfide dehydrogenase associated with the copper-binding protein CopC from the haloalkaliphilic sulfur-oxidizing bacterium Thioalkalivibrio paradoxus ARh 1. *Acta Crystallogr. Sect. D Struct. Biol*. 2018;74(Pt. 7):632-642. doi 10.1107/S2059798318005648

doi 10.18699/bgrs2024-3.1-23

Запросы к базам данных для предсказания и анализа пространственной структуры фермента протеин-дисульфидизомеразы семейства А4 человека и животных

Сушков Р.В.[1]*, Мочалов А.В.[1], Орлов Ю.Л.[2, 3]**, Савина Е.А.[2]

[1] *ФГБОУ ВО ВолгГМУ Минздрава России, Волгоград, Россия*

[2] *Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия*

[3] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

** roman.sushkov.01@mail.ru; ** orlov@bionet.nsc.ru*

Ключевые слова: медицинская информатика; базы данных; протеин-дисульфидизомераза; статистическая обработка; корреляция белковых структур; филогенез

Мотивация и цель: Запросы к биоинформационным базам данных активно используются для предсказания и анализа пространственной структуры белков и ее изменчивости в медицине и генетике. Инструменты биоинформатики помогают анализировать генетические данные в симуляции и моделировании ДНК, РНК и белковых структур, в оценке молекулярных взаимодействий, разработке лекарственных веществ [1]. Протеин-дисульфидизомераза семейства А4 (PDIA4) – изомераза, катализирует образование дисульфидных связей и контролирует перестройки молекул в ЭПР, что обеспечивает их «гибкость», формирует шаперон-опосредованную укладку белка [2]. Нарушение вышеописанных функций приводит к различным дегенеративным изменениям нервной системы, например, болезни Альцгеймера и Паркинсона [3]. Практическая цель данной работы – поиск наиболее схожих по пространственной структуре и функциональной активности белков PDIA4 человека и животных, что может использоваться в медицине для создания лекарственных препаратов на базе его аналогов и для дальнейшего исследования биохимических свойств биомолекулы в лаборатории.
Методы и алгоритмы: Составление списков белков выполнялось на основе запросов к базам данных UniProtKB (https://uniprot.org/) и BioGraph (http://andromeda.matf.bg.ac.rs:54321/) [4]. В результате был получен список из 28 животных белков, анализ структуры которых мог дать позитивный результат в исследовании. Обработка аминокислотных последовательностей выполнялась в формате FASTA. Проводили попарное выравнивание аминокислотных последовательностей PDIA4 человека и животных в программе Clustal Omega (ebi.ac.uk/jdispatcher/msa/clustalo). Полученные вероятности различия двух последовательностей заносились в таблицу Excel «Анализ сходства первичных последовательностей фермента PDIA4 у человека к этому же ферменту других организмов». Параллельно производилось попарное сравнение аминокислотных последовательностей в системе BLAST. Полученными значениями (Total score; Query cover; E-value; Indent; Positives) дополнили таблицу Excel. Данные были обработаны с помощью ранговых оценок, на основании которых получена сводная таблица, включающая в себя три группы белков: наиболее сходные с белком человека, минимально сходные и медиана. Далее с использованием сайта RCSB

(https://www.rcsb.org/) построили 3D модели белков PDIA4. Сравнивались аминокислоты, подверженные модификациям у человеческого и животного белка.
Результаты: В качестве объектов для сравнения использовали белок человека и 28 животных. В результате выравнивания аминокислотных последовательностей в системе Clustal Omega и BLAST было выявлено следующее: наиболее сходен с человеческим вариантом PDIA4 белок хомяка; для проведения исследований *in vivo* следует использовать хомяка или крысу, поскольку они являются стандартным лабораторным животным; для получения чистого фермента и проведения исследований *in vitro* рациональнее рассматривать свинью, ввиду доступности ее тканей. Сравнение 3D структур белка PDIA4 человека и крысы на сайте RCSB дало результат в 0,88 TM-store, высокий процент корреляции.
Выводы: Путем использования запросов к базам данных проведено исследование пространственной структурны белка PDIA4. Произведен поиск первичных аминокислотных последовательностей. Выполнено попарное сравнение белков человека и животных. Получены данные ранговой обработки результатов исследования и построены доверительные интервалы. Найдены и коррелированы 3D модели наиболее близких по структуре белков. Можно сделать вывод о перспективности области исследований по поиску и разработке лекарственных препаратов на матрице белка животных PDIA4 [5], в частности крысы и хомяка, которые показали наибольшую связь с человеческим.

Database queries for prediction and analysis of the spatial structure of the protein disulfide isomerase enzyme of the A4 family of humans and animals

Sushkov R.V.[1]*, Mochalov A.V.[1], Orlov Y.L.[2, 3]**, Savina E.A.[2]

[1] *Volgograd State Medical University, Volgograd, Russia*

[2] *Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

** roman.sushkov.01@mail.ru; ** orlov@bionet.nsc.ru*

Key words: medical informatics; databases; protein disulfide isomerase; statistical processing; correlation of protein structures; phylogeny

Motivation and Aim: Database queries are actively used to predict and analyze the spatial structure of proteins and its variability in medicine and genetics. Bioinformatics tools help to analyze genetic data in the simulation and modeling of DNA, RNA and protein structures, in the assessment of molecular interactions, and in the development of drugs [1]. Protein disulfide isomerase of the A4 family (PDIA4) isomerase, catalyzes the formation of disulfide bonds and controls rearrangements of molecules in the EPR, which ensures their “flexibility”, forms chaperone-mediated protein folding [2]. Violation of the above-described functions leads to various degenerative changes in the nervous system, for example, Alzheimer's and Parkinson's diseases [3]. The practical purpose of this work is to search for the most similar in spatial structure and functional activity of human and animal PDIA4 proteins, which can be used in medicine to create

drugs based on its synthetic analogues, and for further research of the biochemical properties of biomolecules in laboratories.

Methods and Algorithms: Protein lists were compiled based on queries to UniProtKB databases (https://uniprot.org/) and BioGraph (http://andromeda.matf.bg.ac.rs:54321/) [4]. As a result, a list of 28 animal proteins was obtained, the analysis of the structure of which could give a positive result in the study. The processing of amino acid sequences was performed in the FASTA format. Pairwise alignment of human and animal PDIA4 amino acid sequences was performed in the Clustal Omega program (ebi.ac.uk/jdispatcher/msa/clustalo). The obtained probabilities of differences between the two sequences were entered into an Excel spreadsheet "Analysis of the similarity of the primary sequences of the PDIA4 enzyme in humans to the same enzyme of other organisms". In parallel with the above, a pairwise comparison of amino acid sequences in the BLAST system was performed. The obtained values (Total score; Query cover; E-value; Indent; Positives) were added to the Excel table. The data were processed using rank estimates, on the basis of which a summary table was obtained, including three groups of proteins: the most similar to human protein, the least similar and the median. Further using the RCSB website (https://www.rcsb.org/) we built 3D models of PDIA4 proteins. Amino acids subject to modifications in human and animal proteins were compared.

Results: Human protein and 28 animals were used as objects for comparison. As a result of the alignment of amino acid sequences in the Clustal Omega and BLAST system, the following was revealed: the hamster protein is most similar to the human variant of PDIA4; for *in vivo* studies, a hamster or rat should be used, since they are a standard laboratory animal; for obtaining a pure enzyme and conducting *in vitro* studies, it is more rational to consider a pig, due to the accessibility of her tissues. A comparison of the 3D structures of the human and rat PDIA4 protein on the RCSB website yielded a result of 0.88 TM-store, a high percentage of correlation.

Conclusion: By using database queries, a study of the spatial structure of the PDIA4 protein was conducted. A search for primary amino acid sequences has been performed. A pairwise comparison of human and animal proteins was performed. The data of the rank processing of the research results were obtained, and confidence intervals were constructed. 3D models of the proteins closest in structure have been found and correlated. It can be concluded that the field of research on the search and development of drugs based on the animal protein matrix PDIA4 is promising [5]. In particular, rats and hamsters, which showed the greatest connection with the human.

Список литературы/References

1. Mullowney M.W., Duncan K.R., Elsayed S.S. et al. Artificial intelligence for natural product drug discovery. *Nat Rev Drug Discov.* 2023;22(11):895-916. doi 10.1038/s41573-023-00774-7
2. Fu J., Gao J., Liang Z., Yang D. PDI-Regulated Disulfide Bond Formation in Protein Folding and Biomolecular Assembly. *Molecules*. 2020;26(1):171. doi 10.3390/molecules26010171
3. Medinas D.B., Rozas P., Hetz C. Critical roles of protein disulfide isomerases in balancing proteostasis in the nervous system. *J Biol Chem.* 2022;298(7):102087. doi 10.1016/j.jbc.2022.102087
4. Veljković A.N., Orlov Y.L., Mitić N.S. BioGraph: Data Model for Linking and Querying Diverse Biological Metadata. *Int J Mol Sci.* 2023;24(8):6954. doi 10.3390/ijms24086954
5. Kuo T.F., Hsu S.W., Huang S.H. et al. Pdia4 regulates β-cell pathogenesis in diabetes: molecular mechanism and targeted therapy. *EMBO Mol Med.* 2021;13(10):e11668. doi 10.15252/emmm.201911668

doi 10.18699/bgrs2024-3.1-24

Особенности влияния графена и оксида графена на стрессоустойчивость клеток *Escherichia coli*, динамику бактериальной ДНК и ДНК-связывающего белка Dps

Терешкин Э.В.[1]*, Лойко Н.Г.[2], Потокина В.В.[2], Коваленко В.В.[1], Крупянский Ю.Ф.[1], Терешкина К.Б.[1]

[1] *Федеральный исследовательский центр химической физики им. Н.Н. Семенова РАН, Москва, Россия*

[2] *Федеральный исследовательский центр «Фундаментальные основы биотехнологии» РАН, Москва, Россия*

* *ramm@mail.ru*

Ключевые слова: бактериальная жизнеспособность и стрессоустойчивость; стабилизация ДНК бактерий; белок Dps; ДНК-белковые комплексы; молекулярная динамика; моделирование биологических молекул на графене

Мотивация и цель: Исследовать влияние графена и оксида графена на стрессоустойчивость бактерий *Escherichia coli* (*E. coli*). Выявить различия в динамике дезоксирибонуклеиновой кислоты (ДНК) и стабилизирующего ее белка на поверхности графена, оксида графена и в растворе. Полученные данные представляют практический интерес для разработки новых способов стерилизации и для исследования структуры биологических молекул и их комплексов на поверхности графеновых подложек. Результаты исследований могут быть использованы при создании основанных на биологических молекулах наноматериалов с заданными свойствами [1, 2].

Методы и алгоритмы: Бактерии *E. coli* K12 выращивали на среде LB в присутствии разных концентраций графена (Г) и оксида графена (ГО) при температуре 28 °C. Бактериальные популяции, полученные в опытных и контрольных (без добавок Г и ГО) условиях, подвергали 30-минутному воздействию температуры 55 °C и 1–2-минутному воздействию УФ-излучения, определяя количество выживших клеток и оценивая стрессоустойчивость опытных популяций по сравнению с контрольной. Методами молекулярного моделирования исследованы полноатомные модели систем графена, оксида графена, ДНК-связывающего белка Dps бактерии *E. coli* и участков ДНК (25–135 пар нуклеотидов) в несвязанном состоянии и гистоноподобных комплексах. Расчеты молекулярной динамики проводились в полноатомном приближении в периодических ячейках с использованием программного комплекса Gromacs при постоянной температуре 28 °C (поддерживалась с помощью термостата Ланжевена) и давлении 1 атм (поддерживалось с помощью баростата Парринелло–Рамана). Для определения термодинамических характеристик связывания ДНК с молекулами белка Dps применен метод поиска линейной энергии взаимодействия (LIE).

Результаты: Микробиологические исследования показали, что рост бактерий *E. coli* K12 в присутствии Г и ГО влиял на жизнеспособность и стрессоустойчивость выросших популяций. В опытных популяциях после

стрессовых воздействий (55 °C, 30 мин и УФ-облучение, 1–2 мин) количество выживших клеток снижалось на 4–6 порядков, тогда как в контрольной – на 1–3. Эффект был прямо пропорционален количеству Г и ГО, присутствующих в среде во время роста бактерий. При определенных стрессовых режимах клетки опытных популяций погибали полностью, тогда как в контрольных популяциях выживали на уровне 10^6 клеток/мл.

Методами молекулярного моделирования проведено сравнительное исследование конформационной подвижности биологических молекул ферритиноподобного ДНК-связывающего белка Dps (DNA-binding protein from starved cells) *E. coli* и участков ДНК разной длины в растворе, а также на поверхности графена и оксида графена. В многочисленных исследованиях установлено, что Dps является основным белком, обеспечивающим стабилизацию ДНК и выживаемость бактериальных клеток в стрессовых условиях [3]. Он представляет собой гомододекамер с полостью внутри, обладает функциями секвестрации железа в клетке бактерий и обеспечивает выживание бактериальных клеток благодаря способности сокристаллизации с ДНК [3, 4], обеспечивая защиту бактериального генома в стрессовых условиях. Кроме того, это перспективный биополимер для использования в качестве наноматериала [1, 2]. Расчеты молекулярной динамики на 0.1–0.5 мкс траекториях показывают, что графеновые подложки обоих типов существенным образом влияют на подвижность N-концевых (ДНК-связывающих) участков ближайших к поверхности субъединиц белка, связывая их на своей поверхности. Структура белка в целом значительнее стабильнее на подложке из неокисленного графена, чем окисленного, на которой начинает изменяться четвертичная структура белка. Проведенный анализ главных компонент (рис. 1) показывает, что адсорбированные на поверхности неокисленного графена N-концы субъединиц, расположенных вдоль нормали к поверхности подложки, испытывают сильные флуктуации, которые передаются всей субъединице в целом, что приводит к структурным перестройкам внутри молекулы белка (поры расширяются, субъединицы смещаются друг относительно друга).

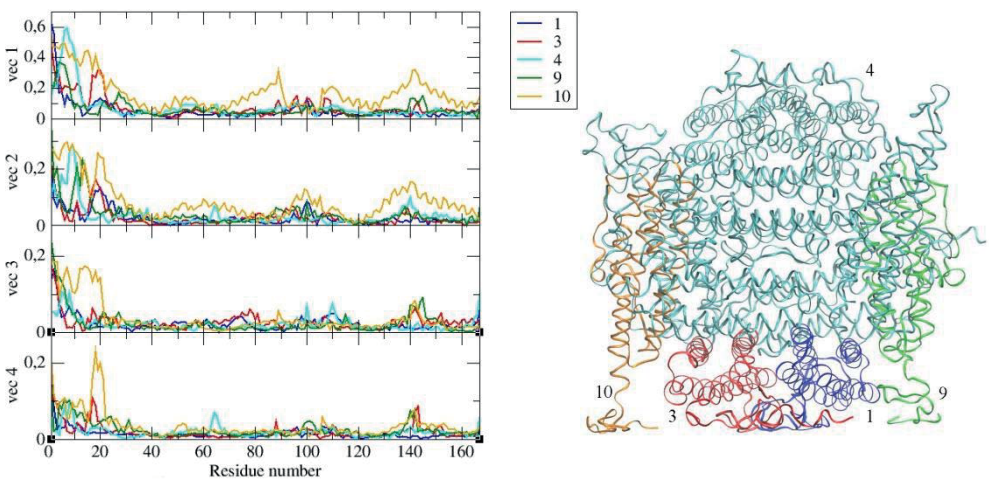


Рис. 1. Слева – анализ среднеквадратичных флуктуаций аминокислотных остатков пяти из 12 субъединиц белка Dps, адсорбированного на поверхность неокисленного графена, относительно четырех собственных векторов (vec1-vec4). Справа – структура белка Dps, цифрами обозначены пять субъединиц, соответствующие графику слева

Исследовалась динамика комплексов Dps-ДНК на графеновых подложках. Показано, что при приближении комплекса к подложке олигонуклеотиды (25 пар нуклеотидов) отрываются от белка и адсорбируются на графене и оксиде графена. Гистоноподобные комплексы Dps с ДНК (165 пар нуклеотидов) также претерпевают структурные перестройки: ДНК начинает отрываться от белка и приближаться к графеновой подложке (рис. 2).

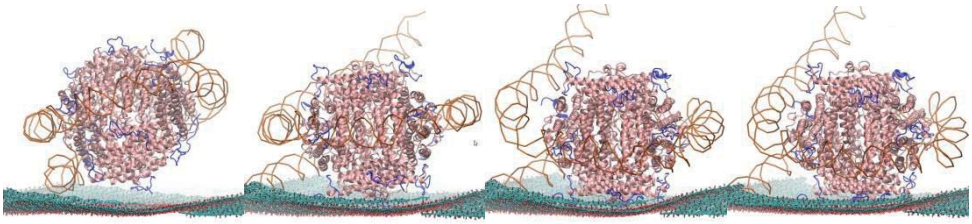


Рис. 2. Комплекс белка Dps с ДНК (165 пар нуклеотидов) на оксиде графена в течение времени траектории (слева направо): начальное положение, 100 нс, 200 нс, 300 нс. Белок Dps – розовый, N-концевые аминокислотные остатки – синие, ДНК – оранжевая спираль, оксид графена – голубая подложка

Образование комплексов ДНК-Dps на графеновых подложках происходит, однако такие комплексы оказываются менее стабильны, чем комплексы в растворе. Свободная энергия связывания ДНК ниже в том случае, если молекула оказывается связанной с графеновой подложкой и белком, нежели только с белком.

Выводы: Обнаружено, что рост бактерий *E. coli* в присутствии Г и ГО приводит к резкому снижению их стрессоустойчивости к температурному и УФ воздействию, что может быть использовано для разработки новых способов стерилизации. В результате проведенных молекулярно-динамических расчетов выявлено влияние подложек графена и оксида графена на динамику и структуру комплексов ДНК-стабилизирующего белка Dps *E. coli* с ДНК. Показано, что неокисленный графен может привести к изменению четвертичной структуры белка. ДНК в комплексах стремится связаться одновременно и с белком, и с подложкой, что приводит к структурным перестройкам или разрушению комплексов Dps-ДНК.

Финансирование: Расчеты проводились на высокопроизводительной вычислительной системе МВС–10П в Межведомственном суперкомпьютерном центре Российской академии наук (МСЦ РАН). Работа выполнена в рамках государственного задания Минобрнауки России (Тема FFZE-2022-0011, № 122040400089-6, 122040800164-6).

Peculiarities of the influence of graphene and graphene oxide on the stress resistance of *Escherichia coli* cells, the dynamics of bacterial DNA and the DNA-binding protein Dps

Tereshkin E.V.[1]*, Loiko N.G.[2], Potokina V.V.[2], Kovalenko V.V.[1], Krupyanskii Y.F.[1], Tereshkina K.B.[1]

[1] *Semenov Federal Research Center for Chemical Physics, RAS, Moscow, Russia*
[2] *Federal Research Center "Fundamentals of Biotechnology", RAS, Moscow, Russia*
* *ramm@mail.ru*

Key words: bacterial viability and stress resistance; stabilization of bacterial DNA; Dps protein; DNA-protein complexes; molecular dynamics; modeling of biological molecules on graphene

Motivation and Aim: To study the effect of graphene and graphene oxide on the stress resistance of *Escherichia coli (E. coli*) bacteria. Identify differences in the dynamics of deoxyribonucleic acid (DNA) and the protein that stabilizes it on the surface of graphene, graphene oxide and in solution. The data obtained are of practical interest for the development of new sterilization methods and for researchers of the structure of biological molecules and their complexes on the surface of graphene substrates. The research results can be used to create nanomaterials based on biological molecules with specified properties [1, 2].
Methods and Algorithms: *E. coli* K12 bacteria were grown in LB medium in the presence of different concentrations of graphene (G) and graphene oxide (GO) at a temperature of 28 °C. Bacterial populations obtained under experimental and control (without G and GO additives) conditions were subjected to 30 min exposure to a temperature of 55 °C and 1–2 min exposure to UV radiation, determining the number of surviving cells and assessing the stress resistance of the experimental populations compared to the control. Using molecular modeling methods, all-atom models of graphene, graphene oxide, the DNA-binding protein Dps of the bacterium *E. coli* and DNA oligonucleotides (25–135 base pairs) in an unbound state and in histone-like complexes were studied. Molecular dynamics simulations were carried out in the all-atom approximation in periodic boxes using Gromacs at a constant temperature of 28 °C (maintained using a Langevin thermostat) and a pressure of 1 atm (maintained using a Parrinello-Raman barostat). To determine the thermodynamic characteristics of DNA binding to Dps protein molecules, the linear interaction energy (LIE) method was used.
Results: Microbiological studies showed that the growth of *E. coli* K12 bacteria in the presence of G and GO influenced the viability and stress resistance of the grown populations. In the experimental populations, after stress exposure (55 °C, 30 min and UV irradiation, 1–2 min), the number of surviving cells decreased by 4 to 6 orders of magnitude, while in the control population by 1 to 3. The effect was directly proportional to the amount of G and GO present in the medium during bacterial growth. Under certain stress conditions, the cells of the experimental populations died completely, while in the control populations they survived at a level of 10^6 cells/ml.
Using molecular modeling methods, a comparative study of the conformational mobility of biological molecules of the ferritin-like DNA-binding protein Dps (DNA-binding protein from starved cells) of *E. coli* and DNA sections of different lengths in solution, as well as on the surface of graphene and graphene oxide, was carried out. Numerous studies have established that Dps is the main protein that ensures DNA stabilization and the survival of bacterial cells under stressful conditions [3]. It is a homododecamer with a cavity inside and has the functions of sequestering iron in the bacterial cell and ensures the survival of bacterial cells due to the ability to co-crystallize with DNA [3, 4], providing protection of the bacterial genome under stressful conditions. In addition, it is a promising biopolymer for use as a nanomaterial [1, 2]. Molecular dynamics simulations on 0.1–0.5 μs trajectories show that graphene substrates of both types

significantly affect the mobility of the N-terminal (DNA-binding) regions of the protein subunits closest to the surface, binding them to their surface. The protein structure as a whole is much more stable on a substrate made of unoxidized graphene than on an oxidized one, on which the quaternary structure of the protein begins to change. The analysis of the principal components (Fig. 1) shows that the N-termini of subunits adsorbed on the surface of unoxidized graphene, located along the normal to the surface of the substrate, experience strong fluctuations, which are transmitted to the entire subunit as a whole, which leads to structural rearrangements inside the protein molecule (pores expand, subunits move relative to each other).

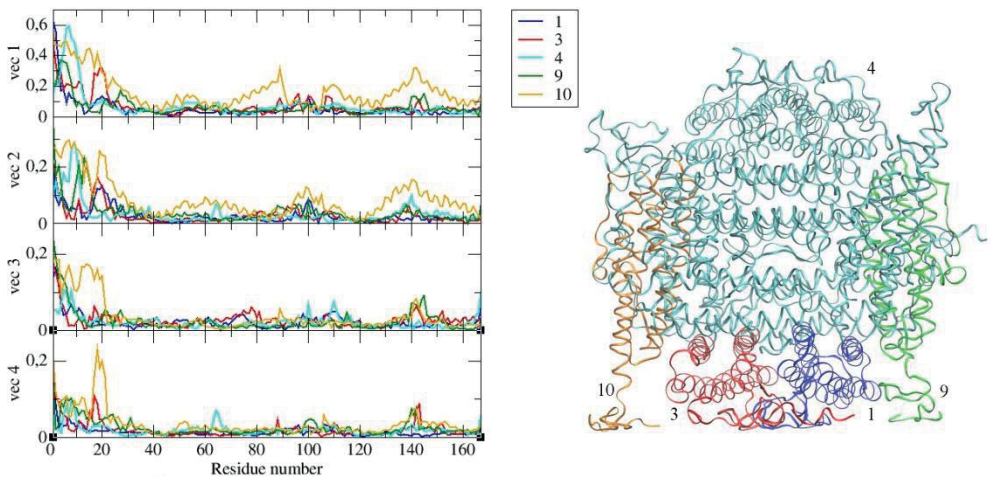


Fig. 1. On the left is an analysis of the root mean square fluctuations of amino acid residues of five of the 12 subunits of the Dps protein adsorbed on the surface of unoxidized graphene, relative to four eigenvectors (vec1-vec4). On the right is the structure of the Dps protein, with numbers indicating the five subunits that correspond to the graphs on the left

The dynamics of Dps-DNA complexes on graphene substrates was studied. It was shown that when the complex approaches the substrate, oligonucleotides (25 nucleotide pairs) are detached from the protein and adsorbed on graphene and graphene oxide. Histone-like complexes of Dps with DNA (165 base pairs) also undergo structural rearrangements: DNA begins to detach from the protein and approaches the graphene substrate (Fig. 2).

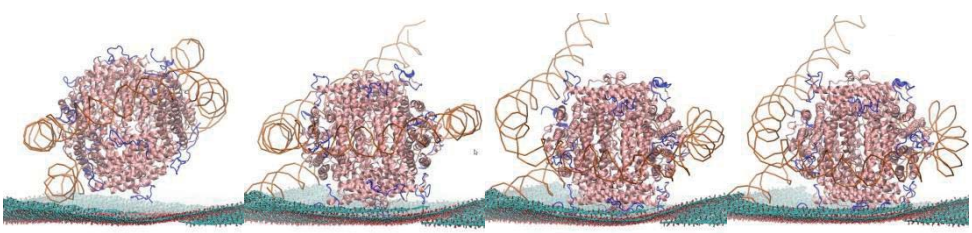


Fig. 2. Dps protein complex with DNA (165 nucleotide pairs) on graphene oxide during the trajectory time (from left to right): initial position, 100 ns, 200 ns, 300 ns. Dps protein – pink, N-terminal amino acid residues – blue, DNA – orange helix, graphene oxide – blue substrate

The formation of DNA-Dps complexes on graphene substrates occurs, but such complexes turn out to be less stable than complexes in solution. The free energy of DNA binding turns out to be lower if it is bound to both the graphene substrate and the protein, rather than only to the protein.

Conclusion: It was found that the growth of *E. coli* bacteria in the presence of G and GO leads to a sharp decrease in their stress resistance to temperature and UV exposure, which can be used to develop new sterilization methods. As a result of molecular dynamics simulations, the influence of G and GO substrates on the dynamics and structure of complexes of the DNA-stabilizing protein Dps of *E. coli* with DNA was revealed. It has been shown that G can lead to changes in the quaternary structure of the protein. DNA in complexes tends to contact both the protein and the substrate (G or GO) simultaneously, which leads to structural rearrangements or destruction of Dps-DNA complexes.

Funding: This study was carried out as part of a state task of the Ministry of Education and Science of Russia on the topic FFZE-2022-0011 (registration numbers 122040400089-6 and 122040800164-6). The computations were performed on MVS-10P at the Joint Supercomputer Center of the Russian Academy of Sciences.

Список литературы/References

1. Kamitake H., Uenuma M., Okamoto N., Horita M., Ishikawa Y., Yamashita I., Uraoka Y. Floating gate memory with charge storage dots array formed by Dps protein modified with site-specific binding peptides. *Nanotechnology*. 2015;26(19):195201. doi 10.1088/0957-4484/26/19/195201
2. Gupta N.K., Karuppannan S.K., Pasula R.R., Vilan A., Martin J., Xu W., May E.M., Pike A.R., Astier H.P.A.G., Salim T., Lim S., Nijhuis C.A. Temperature-Dependent Coherent Tunneling across Graphene-Ferritin Biomolecular Junctions. *ACS Appl Mater Interfaces*. 2022;14(39):44665-44675. doi 10.1021/acsami.2c11263
3. Grant R.A., Filman D.J., Finkel S.E., Kolter R., Hogle J.M. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol.* 1998;5(4):294-303. doi 10.1038/nsb0498-294
4. Tereshkin E.V., Tereshkina K.B., Krupyanskii Yu.F. Predicting Binding Free Energies for Dps Protein-DNA complexes and Crystals Using Molecular Dynamics. *Supercomput Front Innovations*. 2022;9(2):33-45. doi 10.14529/jsfi220203

doi 10.18699/bgrs2024-3.1-25

Динамика связывания ДНК со стресс-ассоциированными ДНК-связывающими бактериальными белками

Терешкина К.Б.[1]*, Терешкин Э.В.[1], Коваленко В.В.[1], Крупянский Ю.Ф.[1], Лойко Н.Г.[2]

[1] *Федеральный исследовательский центр химической физики им. Н.Н. Семенова РАН, Москва, Россия*

[2] *Федеральный исследовательский центр «Фундаментальные основы биотехнологии» РАН, Москва, Россия*

* *quebra-mola@yandex.ru*

Ключевые слова: стабилизация ДНК бактерий; белок Dps; кристаллы ДНК-Dps; ферритиноподобные белки у бактерий; полноатомное молекулярное моделирование

Мотивация и цель: Переход бактериальных клеток в покоящееся состояние при стрессах сопровождается значительным снижением их чувствительности к антибиотикам, антисептикам, дезинфектантам и консервантам, что создает существенные проблемы в области антибактериальных мероприятий для медицины и производств, где требуется соблюдение условий стерильности [1]. Важную роль в сохранении генетического материала бактерий в таких состояниях играют ферритиноподобные ДНК-связывающие белки Dps и их гомологи [2]. Целью данной работы было выявление множественности этих белков среди бактерий, в том числе болезнетворных, и исследование возможностей связывания ими ДНК у ряда бактерий.

Методы и алгоритмы: Методами молекулярного моделирования и биоинформатики проведены сравнительные исследования структур белков Dps и гомологичных им у ряда бактерий и их комплексов с ДНК. Изучение образования и эволюции комплексов проведено методом классической молекулярной динамики на основе программного комплекса Gromacs в полноатомном приближении на траекториях до 1 мкс согласно разработанным ранее протоколам [3].

Результаты: Dps (DNA protection during starvation protein) представляет собой гомододекамер и у бактерии *Escherichia coli* (*E. coli*) выполняет важнейшие функции [2]. Прежде всего, он защищает клетку от опасных ионов Fe^{2+}, аккумулируя их во внутренней полости в связанном состоянии Fe^{3+}---белок. Также, наряду с не содержащим гема бактериальным ферритином FtnA и бактериоферритином Bfr, он связывает бактериальную ДНК, защищая ее в неблагоприятных условиях. Последнее возможно благодаря связыванию ДНК свободными N-концами белка и способности в определенных условиях легко образовывать сокристаллы ДНК-Dps. Изучение литературных источников и баз данных дает понимание того, что гомологичные Dps *E. coli* белки распространены повсеместно у бактерий и архей. В базе данных Uniprot обнаруживается более 25 тыс. записей. Несколько десятков расшифрованных экспериментальными методами структур, относящихся к 37 видам бактерий и архей, выложены в базе данных PDB.

Как хорошо видно из результатов выравнивания первичных последовательностей гомологов Dps некоторых бактерий (рис. 1), аминокислотные остатки, входящие в ферроксидазные центры и отвечающие за связывание Fe^{2+}, у всех бактерий совпадают. В то время как общее сходство последовательностей ограничивается 20 %. Основные различия в ДНК-связывающей способности должны определяться структурой N- и C-концевых «хвостов», которые захватывают ДНК и обеспечивают начальные этапы ее упаковки.

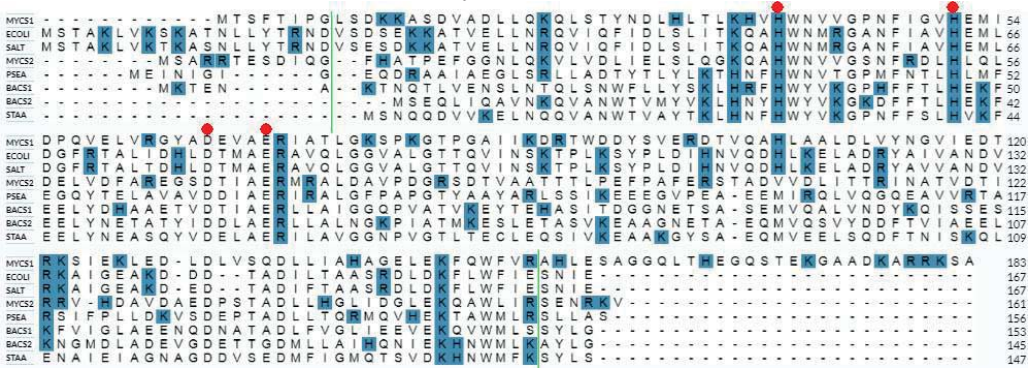


Рис. 1. Выравнивание первичных аминокислотных последовательностей ДНК-связывающих белков бактерий (сверху вниз): *Mycobacterium smegmatis* Dps1, *Escherichia coli* Dps, *Salmonella typhimurium* Dps, *Mycobacterium smegmatis* Dps2, *Pseudomonas aeruginosa* Dps, *Bacillus subtilis* MrgA, *Bacillus subtilis* Dps, *Staphylococcus aureus* MrgA. Выделены положительно заряженные аминокислотные остатки. Красными кружками обозначены аминокислотные остатки, входящие в сайты связывания Fe^{2+}. Зелеными линиями примерно ограничены N-концевые и C-концевые участки

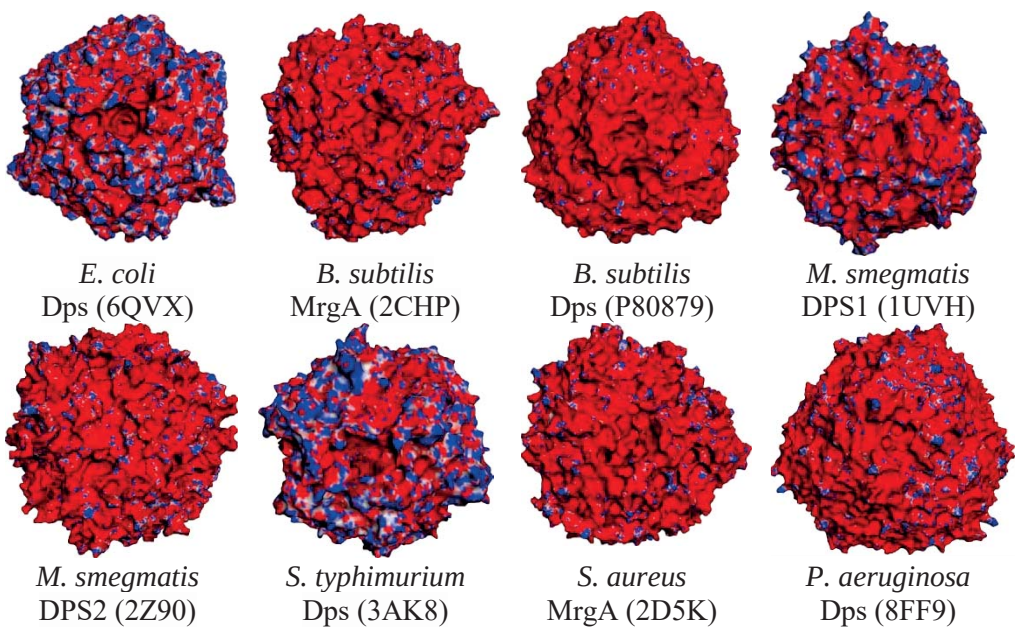


Рис. 2. Поверхности Ван-дер-Ваальса белков Dps, окрашенные в соответствии с рассчитанными электростатическими потенциалами от электроотрицательного (красный) до электроположительного (синий) при pH = 7.2. Приведен вид со стороны ферритиновой поры. Под рисунками указаны названия бактерий, белков, код PDB (код AlphaFold для Dps *Bacillus subtilis*)

Следует отметить, что в процессе образования ДНК-белкового комплекса ДНК и белок подстраиваются друг под друга. ДНК изгибается. Радиусы гирации белков при сближении с ДНК (рис. 3) снижаются по сравнению с белками в растворе. Способность образования рассматриваемых комплексов имеет температурную и pH зависимость.

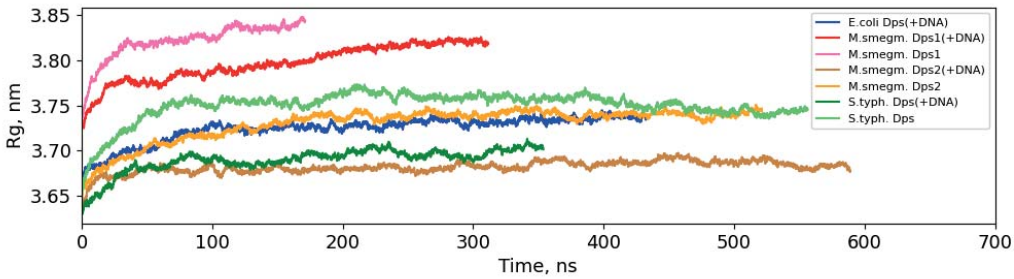


Рис. 3. Радиусы гирации белков Dps в растворе и в комплексах с ДНК

Выводы: В работе исследован ряд ДНК-связывающих ферритиноподобных бактериальных белков Dps. Выравнивание первичных последовательностей указывает на высокую консервативность связывающих железо центров и широкое разнообразие длин, а также последовательностей концевых участков. Показано, что способные связывать ДНК белки встречаются и у грамположительных, и у грамотрицательных бактерий. Основное влияние на способность связывать ДНК оказывает заряд поверхности белка и наличие длинных (порядка 10–20 аминокислотных остатков) свободных концов, которые встраиваются в бороздки ДНК и подтягивают ДНК к поверхности белка.

Финансирование: Работа выполнена при поддержке гранта РНФ № 23-24-00250. Расчеты проводились на высокопроизводительной вычислительной системе МВС–10П в Межведомственном суперкомпьютерном центре Российской академии наук (МСЦ РАН).

Dynamics of DNA binding to stress-associated DNA-stabilizing bacterial proteins

Tereshkina K.B.[1]*, Tereshkin E.V.[1], Kovalenko V.V.[1], Krupyanskii Y.F.[1], Loiko N.G.[2]

[1] *Semenov Federal Research Center for Chemical Physics, RAS, Moscow, Russia*

[2] *Federal Research Center "Fundamentals of Biotechnology", RAS, Moscow, Russia*

* *quebra-mola@yandex.ru*

Key words: bacterial DNA stabilization; Dps protein; DNA-Dps crystals; ferritin-like proteins in bacteria; all-atom molecular modeling

Motivation and Aim: The transition of bacterial cells to a starved state under stress is accompanied by a significant decrease in their sensitivity to antibiotics, antiseptics, disinfectants and preservatives, which creates significant problems in the field of antibacterial measures for medicine and industries where sterility conditions are required [1]. Ferritin-like DNA-binding proteins Dps and homologues play an important role in

preserving the genetic material of bacteria in such conditions [2]. The purpose of this work was to determine the prevalence of these proteins among bacteria, including pathogenic ones, and to study the possibilities of their DNA binding in a number of bacteria.

Methods and Algorithms: Using molecular modeling and bioinformatics methods, comparative studies of the structures of Dps proteins and homologous proteins in a number of bacteria and their complexes with DNA were carried out. The formation and evolution of complexes was studied using the method of classical molecular dynamics based on the Gromacs software package in the all-atom approximation on trajectories up top to 1 μs according to a previously developed protocols [3].

Results: Dps (DNA protection during starvation protein) is a homododecamer and performs essential functions in the bacterium *Escherichia coli* (*E. coli*) [2]. First, it protects the cell from dangerous Fe^{2+} ions, accumulating them in the internal cavity in the bound state of Fe^{3+}---protein. Also, along with the non-heme bacterial ferritin FtnA and bacterioferritin Bfr, it binds bacterial DNA, protecting it under stress conditions. The latter is possible due to the binding of DNA by the free N-termini of the protein and the ability, under certain conditions, to easily form DNA-Dps co-crystals. A study of literature sources and databases provides an understanding that proteins homologous to *E. coli* Dps are ubiquitous in bacteria and archaea. More than 25 thousand entries are found in the Uniprot database. Several dozen structures related to 37 species of bacteria and archaea, resolved by experimental methods, are posted in the PDB database.

As is clearly seen from the results of alignment of the primary sequences of Dps homologs of some bacteria (Fig. 1), the amino acid residues included in the ferroxidase centers and responsible for Fe^{2+} binding are the same in all bacteria. While the overall sequence similarity is limited to 20 %. The major differences in DNA-binding capacity is apparently determined by the structure of the N- and C-terminal "tails" that capture DNA and provide the initial steps of its packaging.

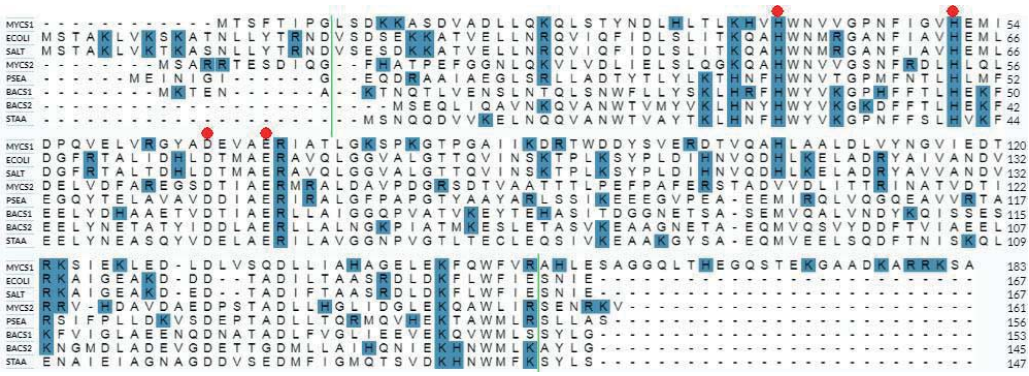


Fig. 1. Alignment of primary amino acid sequences of bacterial DNA-binding proteins (from top to bottom): *Mycobacterium smegmatis* Dps1, *Escherichia coli* Dps, *Salmonella typhimurium* Dps, *Mycobacterium smegmatis* Dps2, *Pseudomonas aeruginosa* Dps, *Bacillus subtilis* MrgA, *Bacillus subtilis* Dps, *Staphylococcus aureus* MrgA. Positively charged amino acid residues are highlighted. Red circles indicate amino acid residues included in Fe^{2+} binding sites. Green lines roughly delimit the N-terminal and C-terminal regions

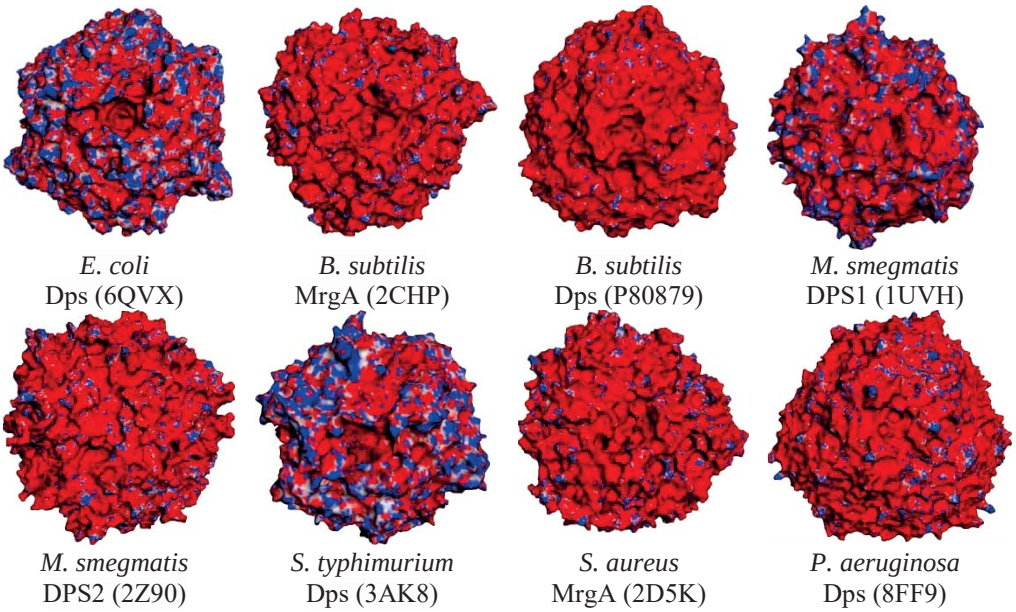


Fig. 2. Van der Waals surfaces of Dps proteins, colored according to calculated electrostatic potentials from electronegative (red) to electropositive (blue) at pH=7.2. The view is shown from the side of the ferritin pore. Below the figures are the names of bacteria, proteins, and the PDB code (AlphaFold code for Dps *Bacillus subtilis*)

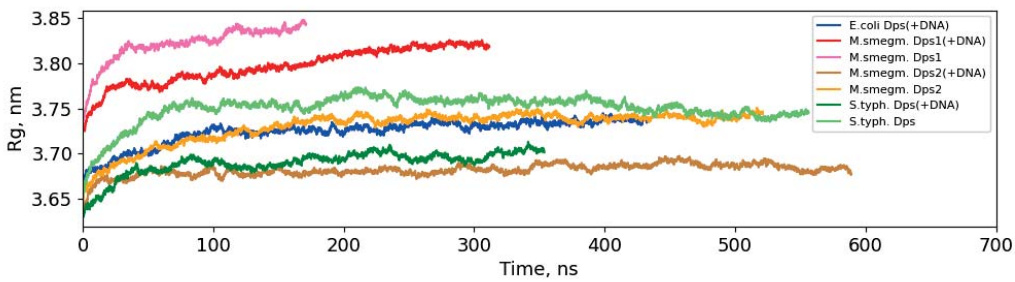


Fig. 3. Radii of gyration for Dps proteins in solution and in complexes with DNA

It should be noted that during the formation of a DNA-protein complex, DNA and protein adapt to each other. DNA bends. The radii of gyration of proteins when approaching DNA (Fig. 3) decrease compared to proteins in solution. The ability to form the complexes under consideration is temperature and pH dependent.

Conclusion: In this work, a number of DNA-binding ferritin-like bacterial proteins Dps were studied. The alignment of the primary sequences indicates high conservation of the iron-binding centers and a wide variety of lengths as well as sequences of the terminal regions. It has been shown that proteins capable of binding DNA are found in both gram-positive and gram-negative bacteria. The main influence on the ability to bind DNA is exerted by the surface charge of the protein and the presence of long (about 10–20 amino

acid residues) free termini, which are embedded in DNA grooves and pull the DNA to the protein surface.

Funding: The research was financially supported by RSF Grant No. 23-24-00250. The computations were performed on MVS-10P at the Joint Supercomputer Center of the Russian Academy of Sciences.

Список литературы/References

1. Windels E.M., Van den Bergh B., Michiels J. Bacteria under antibiotic attack: Different strategies for evolutionary adaptation. *PLoS Pathog.* 2020;16(5):e1008431. doi 10.1371/journal.ppat.1008431
2. Orban K., Finkel S.E. Dps Is a Universally Conserved Dual-Action DNA-Binding and Ferritin Protein. *J Bacteriol*. 2022;17;204(5):e0003622. doi 10.1128/jb.00036-22
3. Tereshkin E.V., Loiko N.G., Tereshkina K.B., Kovalenko V.V., Krupyanskii Y.F., Possible mechanisms of 4-hexylresorcinol influence on DNA and DNA–Dps nanocrystals affecting stress sustainability of Escherichia coli. *Russ J Phys Chem B.* 2022;16(4):726. doi 10.1134/S1990793122040285

doi 10.18699/bgrs2024-3.1-26

Рациональный дизайн терминальной дезоксинуклеотидилтрансферазы человека

Укладов Е.О.[1, 2]*, Тюгашев Т.Е.[1], Кузнецов Н.А.[1, 2]

[1] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

[2] *Факультет естественных наук, Новосибирский государственный университет, Новосибирск, Россия*

* *e.ukladov@g.nsu.ru*

Ключевые слова: полимеразы; терминальная дезоксинуклеотидилтрансфераза; молекулярное моделирование; молекулярная динамика; водородные связи

Мотивация и цель: Терминальная дезоксинуклеотидилтрансфераза (TdT) вместе с полимеразами β, λ, μ и X принадлежит к X-семейству ДНК-полимераз. TdT является уникальной полимеразой для позвоночных и участвует в присоединении случайных нуклеотидов на концы ДНК в процессе V(D)J-рекомбинации [1]. Благодаря уникальной способности к 5'-3'-безматричному синтезу за счет структуры петли 1, TdT нашла применение в молекулярной биологии и биотехнологии [2]. Однако существенным ограничением применения данного фермента является выраженная селективность TdT человека к присоединяемым нуклеотидам, падающая в ряду: dGTP > dCTP ≈ dATP > dTTP [3]. Поэтому целью исследования стало *in silico* моделирование мутантных форм TdT, обладающих измененной селективностью по отношению к dNTP, а также увеличенной каталитической эффективностью.

Методы и алгоритмы: С помощью программы Modeller были получены пре- и посткаталитические комплексы TdT с гексамерным ДНК-праймером на основании кристаллографических структур (PDB ID: 4I27 и 4I29 соответственно). Для параметризации dNTP в прекаталитическом комплексе использовался сервер R.E.D. [4]. Для ионов Mg^{2+} в активном центре (АЦ) была взята dummy-model с разнесенными зарядами. Молекулярная динамика проводилась в ПО GROMACS с силовыми полями 14SB для белка и OL15 для ДНК с моделью воды TIP3P и добавлением ионов Na и Cl в концентрации 0.1 М. Моделирование проводили при температуре 300 K. Результаты анализировали на языке Python с использованием пакетов pandas, matplotlib.pyplot, seaborn и scipy.stats.

Результаты: Исследование включает две основные части: 1) изучение роли ряда аминокислотных остатков в петле 1 и АЦ фермента; 2) предсказание замен, потенциально изменяющих селективность и каталитическую эффективность TdT. На основании филогенетического анализа, кристаллографических структур и моделирования фермента дикого типа для проведения функционального анализа отобраны остатки Asp345, Asp395, Leu397, Phe400, Glu456, Asp472, входящие в состав АЦ или петли 1. Для фермента дикого типа и мутантных форм, содержащих замены данных аминокислотных остатков, были получены модельные структуры двойных белок-ДНК и тройных белок-ДНК-dNTP комплексов.

Полученные данные выявили, что замена D345E, меняющая один из аминокислотных остатков, координирующих каталитические ионы Mg^{2+} в АЦ, в одной из траекторий приводит к стабильному увеличению расстояния между Pα-

фосфата присоединяемого нуклеотида и 3'-концевым кислородом ДНК-праймера. В других траекториях наблюдается выворачивание бокового радикала His342 из АЦ. Известно, что замена его позиционного гомолога H429A в полимеразе μ человека приводит к снижению независимой от матрицы активности [5], а мутантная форма TdT, содержащая замену D345E, имеет на два порядка меньшую по сравнению с диким типом активность [6]. В связи с этим можно заключить, что TdT обладает стабильным активным центром и даже незначительное изменение его геометрии может приводить к уменьшению ферментативной активности.

На следующем этапе было проведено аланиновое сканирование трех аминокислотных остатков Leu397, Phe400 и Asp472. Согласно кристаллографическим структурам, Leu397 «расклинивает» азотистые основания (АО) последнего и предпоследнего нуклеотидов ДНК, нарушая их стэкинг-взаимодействие и стабилизируя стэкинг-взаимодейстие АО последнего и встраиваемого нуклеотидов. При замене L397A описанный эффект исчезает: АО dNTP, последнего и предпоследнего нуклеотидов располагаются параллельно, при этом несколько меняется положение dNTP в АЦ, что, видимо, является каталитически некомпетентным состоянием. Остаток Phe400 участвует в стэкинг-взаимодействии с Phe384 и Trp449 и опосредованно стабилизирует положение АО dNTP. При замене F400A нарушается цепочка стэкинг-взаимодействий, а АО пиримидиновых нуклеотидов (dCTP, dTTP) в 1/3 траекторий принимают перпендикулярное положение по отношению к АО 3'-концевого нуклеотида, что, по-видимому, приводит к каталитически некомпетентному состоянию.

При моделировании TdT дикого типа было обнаружено множество ионных мостиков и водородных связей между петлей 1 и остовом белка. Наиболее стабильными оказались Asp395-Arg457 (образует в среднем 1.5 водородной связи), Asp398-Trp449 (1.3), Gln401-Leu436 (1.5), Ser391-Asp472 (1.6). Предполагалось, что при замене D472A исчезнет один из контактов и положение петли 1 станет менее стабильным. Однако при исчезновении водородной связи Ser391-Asp472 в составе мутантной формы D472A другие связи стали стабильнее: Asp395-Arg457 (2), Asp398-Trp449 (1.7), Gln401-Leu436 (1.9). Таким образом, общая подвижность петли 1 не изменилась при внесении данной замены.

При моделировании комплексов фермента дикого типа было обнаружено, что остатки Asp395 и Glu456 образуют стабильные водородные связи с dGTP (наиболее эффективно встраиваемым нуклеотидом) и практически не образуют с другими dNTP. Было выдвинуто предположение, что данные остатки могут играть ключевую роль в механизме субстратной селективности фермента, и, таким образом, для выравнивания селективности необходимо увеличить количество водородных связей с другими нуклеотидами. Боковые цепи аминокислотных остатков Asp и Glu выступают акцепторами водородной связи, потому были предложены и проанализированы их замены на Asn и Gln, способные выступать и акцепторами, с целью увеличения количества водородных связей между АО dNTP и белком. Среди ортологов TdT в соответствующем Asp395 положении у четверти от общего числа последовательностей представлен Glu. Дополнительно, по аналогии с TdT *Cottoperca gobio*, была выбрана форма D395K+E456Q. Итого были получены модели фермент-субстратных комплексов, содержащих замены: D395E, D395N, D395Q, E456Q, E456N, D395K+E456Q, D395N+E456N, D395Q+E456Q, D395Q+E456N.

Анализ количества водородных связей между белком и АО (рис. 1) показал, что среди одиночных мутантных форм наибольшее количество водородных связей по сравнению с ферментом дикого типа образовывали D395N и E456N, среди двойных – D395N+E456N. Двойная аспарагиновая мутантная форма единственная показала статистически увеличение количества водородных связей относительно дикого типа с тремя нуклеотидами: dATP, dTTP и dCTP.

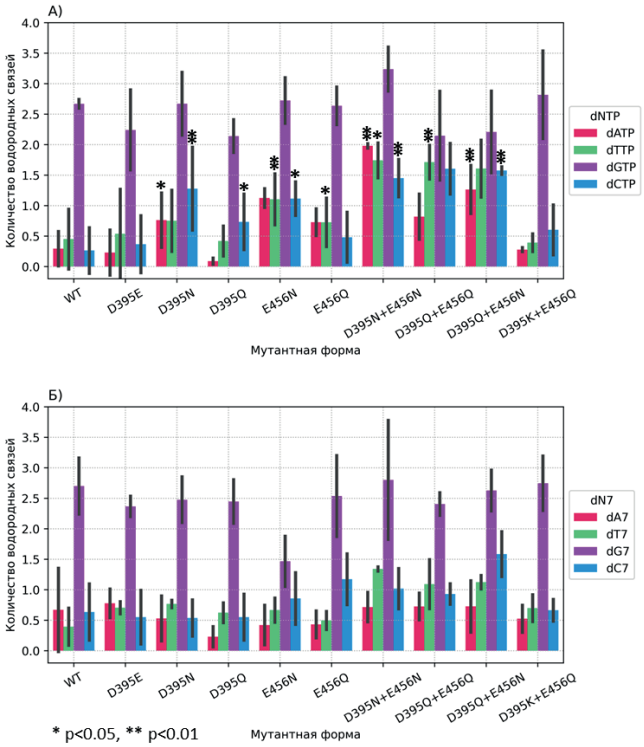


Рис. 1. Среднее количество водородных связей между белком и АО встраиваемого/3'-концевого нуклеотида для пре- (А) и пост-каталитических (Б) комплексов соответственно

Выводы: Выяснена роль аминокислотных остатков Asp345, Asp395, Leu397, Phe400, Glu456, Asp472 в связывании и стабилизации положения dNTP. Предложены аминокислотные замены D395N, E456N и D395N+E456N, потенциально повышающие субстратную селективность к dATP, dCTP и dTTP.
Финансирование: Исследование поддержано грантом РНФ № 21-64-00017.

Rational design of human terminal deoxynucleotidyl transferase

Ukladov E.O.[1, 2]*, Tyugashev T.E.[1], Kuznetsov N.A.[1, 2]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

* *e.ukladov@g.nsu.ru*

Key words: polymerases; terminal deoxynucleotidyltransferase; molecular modelling; molecular dynamics; hydrogen bonds

Motivation and Aim: Terminal deoxynucleotidyl transferase (TdT), along with polymerases β, λ, μ, and X, belongs to the X-family of DNA polymerases. TdT serves as a unique polymerase for vertebrates and participates in the addition of random nucleotides at DNA termini during V(D)J recombination [1]. Due to its distinctive ability for 5'-3' template-independent synthesis facilitated by loop 1 structure, TdT has found applications in molecular biology and biotechnology [2]. However, a significant limitation in the application of this enzyme is the pronounced selectivity of human TdT towards the added nucleotides, ranking as follows: dGTP > dCTP ≈ dATP > dTTP [3]. Therefore, the research objective encompasses *in silico* modeling of mutant forms of TdT, possessing altered selectivity towards dNTPs as well as increased catalytic efficiency.

Methods and Algorithms: Pre- and post-catalytic complexes of TdT with a hexameric DNA primer were obtained using the Modeller program based on crystallographic structures (PDB ID: 4I27 and 4I29 respectively). The parameterization of dNTP in the pre-catalytic complex utilized the R.E.D. server [4]. For Mg^{2+} ions in the active site (AS), a dummy model with distributed charges was employed. Molecular dynamics simulations were conducted using the GROMACS software package with 14SB force fields for the protein and OL15 for DNA, employing the TIP3P water model, and adding Na^{+} and Cl^{-} ions at a concentration of 0.1 M. The simulations were performed at a temperature of 300 K. Results analysis was carried out using the Python language with the utilization of the pandas, matplotlib.pyplot, seaborn, and scipy.stats packages.

Results: This study comprised two main parts: (1) investigating the role of several amino acid residues within loop 1 and the active site (AS) of the enzyme, and (2) predicting substitutions potentially altering the selectivity and catalytic efficiency of TdT.

Based on phylogenetic analysis, crystallographic structures, and wild-type enzyme modeling for functional analysis, residues Asp345, Asp395, Leu397, Phe400, Glu456, and Asp472, which are part of the AS or loop 1, were selected. Model structures of double protein-DNA and triple protein-DNA-dNTP complexes were obtained for both the wild-type enzyme and mutant forms containing substitutions of these amino acid residues.

The obtained data revealed that the substitution D345E, altering one of the amino acid residues coordinating catalytic Mg^{2+} ions in the AS, led to a stable increase in the distance between the Pα phosphate of the incoming nucleotide and the 3'-end oxygen of the DNA primer in one trajectory. In other trajectories, a flipping of the sidechain of His342 from the AS was observed. It is known that the substitution of its positional homolog H429A in human polymerase μ leads to a decrease in template-independent activity [5], and the mutant form of TdT containing the D345E substitution exhibits activity two orders of magnitude lower compared to the wild type [6]. Hence, it can be concluded that TdT possesses a stable active site, and even minor changes in its geometry may lead to decreased enzymatic activity.

In the next stage, alanine scanning of three amino acid residues, Leu397, Phe400, and Asp472, was conducted. According to crystallographic structures, Leu397 “wedges” the nitrogenous bases (NB) of the last and penultimate nucleotides of DNA, disrupting their stacking interaction and stabilizing the stacking interaction between the NB of the last nucleotide and the embedded nucleotide. Upon substitution with L397A, this effect disappears: the NB of dNTP, the last, and penultimate nucleotides align parallelly, with a slight alteration in the position of dNTP in the AS, which appears to be a catalytically incompetent state.

The residue Phe400 participates in stacking interactions with Phe384 and Trp449, indirectly stabilizing the position of dNTP NB. Upon substitution with F400A, the stacking interaction chain is disrupted, and in one-third of the trajectories, the NB of pyrimidine nucleotides (dCTP, dTTP) adopt a perpendicular position with respect to the NB of the 3'-end nucleotide, presumably leading to a catalytically incompetent state.

During the modeling of the wild-type TdT enzyme, numerous ionic bridges and hydrogen bonds between loop 1 and the protein backbone were identified. The most stable ones were found to be Asp395-Arg457 (averaging 1.5 hydrogen bonds), Asp398-Trp449 (1.3), Gln401-Leu436 (1.5), and Ser391-Asp472 (1.6). It was hypothesized that upon substitution D472A, one of the contacts would disappear, rendering the position of loop 1 less stable. However, when the hydrogen bond Ser391-Asp472 vanished in the mutant form D472A, other bonds became more stable: Asp395-Arg457 (2), Asp398-Trp449 (1.7), Gln401-Leu436 (1.9). Thus, the overall mobility of loop 1 remained unchanged following this substitution.

During the modeling of wild-type enzyme complexes, it was discovered that residues Asp395 and Glu456 formed stable hydrogen bonds with dGTP (the most efficiently incorporated nucleotide) and practically did not form bonds with other dNTPs. It was postulated that these residues might play a crucial role in the enzyme's substrate selectivity mechanism, and thus, to equalize selectivity, it was necessary to increase the number of hydrogen bonds with other nucleotides. The side chains of Asp and Glu amino acid residues act as hydrogen bond acceptors, hence substitutions with Asn and Gln, capable of acting as acceptors, were proposed and analyzed to increase the number of hydrogen bonds between dNTP NBs and the protein. Among the TdT orthologs, Glu was present at the corresponding position to Asp395 in a quarter of the total sequences. Additionally, following the analogy with TdT *Cottoperca gobio*, the D395K+E456Q form was selected. As a result, models of enzyme-substrate complexes containing substitutions D395E, D395N, D395Q, E456Q, E456N, D395K+E456Q, D395N+E456N, D395Q+E456Q, D395Q+E456N were obtained.

The analysis of the number of hydrogen bonds between the protein and NBs (Fig. 1) revealed that among the single mutant forms, D395N and E456N formed the highest number of hydrogen bonds compared to the wild-type enzyme. Among the double mutants, D395N+E456N showed the highest increase in the number of hydrogen bonds relative to the wild type with three nucleotides: dATP, dTTP, and dCTP.

Conclusion: The role of the amino acid residues Asp345, Asp395, Leu397, Phe400, Glu456, and Asp472 in binding and stabilizing the position of dNTP has been elucidated. Amino acid substitutions D395N, E456N, and D395N+E456N have been proposed, potentially enhancing substrate selectivity towards dATP, dCTP, and dTTP.

Funding: The study is supported by the grant from the Russian Science Foundation, No. 21-64-00017.

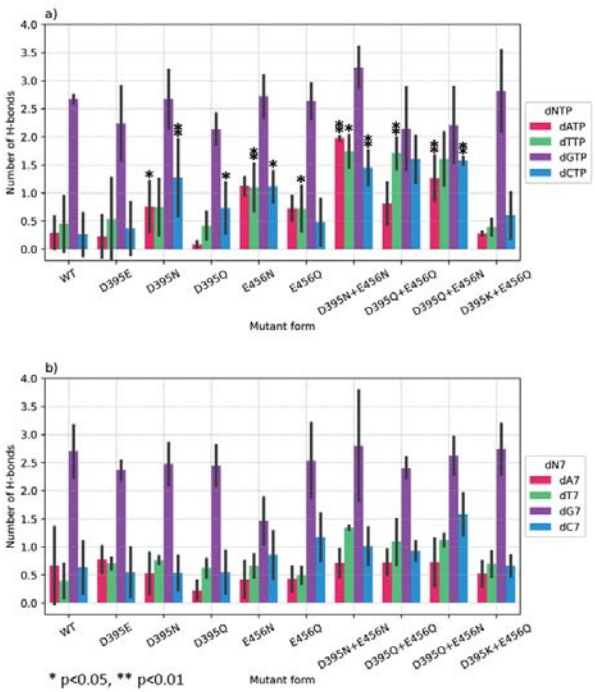


Fig. 1. Mean number of hydrogen bonds between the protein and the NB of the incoming/3'-end nucleotide for pre- (ternary) (A) and post-catalytic (binary) complexes (B) respectively

Список литературы/References

1. Schatz D.G. Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol*. 2011;11(4):251-263
2. Ashley J., Potts I.G., Olorunniji F.J. Applications of Terminal Deoxynucleotidyl Transferase Enzyme in Biotechnology. *ChemBioChem*. 2022;2022:e202200510
3. Kuznetsova A.A., Tyugashev T.E., Alekseeva I.V. et al. Insight into the mechanism of DNA synthesis by human terminal deoxynucleotidyltransferase. *Life Sci Alliance*. 2022;5(12):e202201428
4. Vanquelef E., Simon S., Marquant G. et al. R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res*. 2011;39:W511-W517. doi 10.1093/nar/gkr28
5. Moon A.F., Garcia-Diaz M., Batra V.K. et al. The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair*. 2007;6(12):1709-1725. doi 10.1016/j.dnarep.2007.05.009
6. Yang B., Gathy K.N., Coleman M.S. Mutational analysis of residues in the nucleotide binding domain of human terminal deoxynucleotidyl transferase. *J Biol Chem*. 1994;269(16):11859-11868

doi 10.18699/bgrs2024-3.1-27

О физических основах аминокислотного кода для 3D-структуры белков

Шайтан К.
МГУ им. М.В. Ломоносова, Москва, Россия
shaytan49@yandex.ru

Ключевые слова: линейные полимеры; полипептиды; топография многомерных энергетических поверхностей; динамика фолдинга; эффекты вязкости среды; топологии конфигурационного пространства и симметрии относительно перестановок одинаковых мономерных звеньев; температура денатурации

Мотивация и цель: В настоящее время считается установленным, что пространственная структура белков определяется аминокислотной последовательностью и соответствует минимуму свободной энергии макромолекулы [1]. Имеются некие общие представления о том, что стабильность 3D-структуры требует наличия энергетической щели между нативной конформацией и ближайшими метастабильными конформациями цепи [2], а также о некотором относительно гладком устройстве энергетической поверхности (принцип минимальной фрустрации [3]) для преодоления парадокса Левинталя [4]. Описание современного состояния проблемы можно найти в недавнем обзоре [5]. Однако, конкретные физические аспекты, которые определяют процесс фолдинга долгое время оставались не очень понятными ввиду сложности объекта и отсутствия адекватных математических инструментов для описания данного процесса. Популярность получили компьютерные методы предсказания пространственной структуры белков, которые не основаны на физике фолдинга, а используют технологи ИИ [6]. Эти алгоритмы работают быстро, но с переменным успехом. В основном, хорошо для гомологичных структур. Мы в данном случае исходим из презумпции первенства физической картины и закономерностей для биологических процессов сколь сложными они бы не представлялись на данный момент [7]. Раскрытие физических механизмов организации аминокислотного кода для пространственной структуры белков имеет принципиальное значение для понимания физико-химических основ жизни и способно существенно изменить имеющуюся парадигму исследований в структурной биологии. Развиваемые представления имеют отношение не только к полипептидам, но и к формированию пространственных структур других линейных полимеров и биополимеров.
Методы и алгоритмы: Развиваемые представления основаны на строении линейных полимеров, конформация которых определяется поворотами по двугранным углам вокруг химических связей. Вклад колебаний по валентным степеням свободы в изменение конфигурации макромолекулы является пренебрежимо малым и не рассматривается. Топология конфигурационного пространства линейного полимера соответствует топологии многомерного тора. Функция потенциальной энергии (ПЭ) задана на многомерном торе и представлена в виде разложения в многомерный ряд Фурье. Наличие одинаковых или почти одинаковых мономерных звеньев в длинной полимерной цепи приводит к симметрии поверхности ПЭ относительно перестановки соответствующих

угловых переменных. Это приводит к зависимости коэффициентов разложения от инвариантов алгебраических векторов номеров гармоник, что позволяет моделировать конкретные топографии многомерных энергетических поверхностей и вызванные ими кинетические и термодинамические эффекты. Движение репрезентативной точки (РТ) по многомерной поверхности ПЭ описывается уравнениями механики в вязкой среде. В жидких средах (с вязкостью порядка вязкости воды) инерциальными членами для атомных движений можно пренебречь. Методами многомерной геометрии с учетом асимптотических свойств функций, заданных на многомерных сферах можно установить ряд общих закономерностей для решений системы уравнений движения и правил движения РТ по потенциальной поверхности. Полученные правила движения РТ по многомерной поверхности ПЭ подтверждаются расчетом траекторий молекулярной динамики [7].

Результаты: Можно сформулировать следующие правила движения РТ по многомерной поверхности ПЭ макромолекулы в вязкой среде, которые обеспечивают кинематические связи при фолдинге линейных полимеров.

1. Скорость убыли ПЭ на каждом шаге механической траектории максимальна по сравнению с другими виртуальными перемещениями.
2. Скорость диссипации энергии минимальна по сравнению с другими виртуальными перемещениями. Здесь нет противоречия с п.1 так как утверждения 1 и 2 действуют в разных областях фазового пространства макромолекулы.
3. Скорость диссипации энергии, а также скорость убыли ПЭ макромолекулы практически равномерно "размазаны" по степеням свободы. Это приводит к гладким траекториям для движения РТ с избеганием резких перепадов рельефа ПЭ по относительно небольшому числу степеней свободы (принцип «начинающего горнолыжника»).
4. Как следствие, возникает корреляция конформационных движений линейной полимерной цепи с тенденцией формирования спиральных структур.

Для топографии многомерной поверхности ПЭ линейного полимера с одинаковыми или почти одинаковыми мономерными звеньями возникают следующие закономерности.

1. Поверхность ПЭ для линейных полимеров, построенных из почти одинаковых звеньев, может быть описана в терминах двух обобщенных переменных: суммы квадратов отклонений торсионных углов от точки глобального минимума ПЭ и суммы отклонений углов от точки глобального минимума ПЭ.
2. Топография поверхности ПЭ представляет собой систему вложенных энергетических воронок, разделенных энергетическими барьерами, высота которых убывает по мере удаления от точки глобального минимума. Центральная воронка является самой глубокой.
3. При заданных значениях обобщенных переменных вычисляется энтропия и свободная энергия полимерной цепи. При этом возникает параметр характеристической температуры T_0, который соответствует удельному выигрышу ПЭ при сворачивании цепи, приходящемуся на 1 конформационную степень свободы в мономерном звене. При $T>0.26\ T_0$ свернутое состояние становится неустойчивым. При температуре денатурации порядка 60С в ситуации альфа-аминокислот (2 степени свободы на мономер в основной цепи) выигрыш в потенциальной энергии на 1 мономер должен составлять примерно 5 ккал/моль, что близко к энергии водородной связи.

Выводы: Вышеперечисленные закономерности топографии поверхности ПЭ и правила движения РТ дают логическое объяснение многим весьма разнородным экспериментальным фактам для кинетики и термодинамики процессов фолдинга и рефолдинга [7]. Что касается физических основ возникновения аминокислотного кода для 3D-структуры белков, то мы формулируем следующую гипотезу. На ранней, добиологической стадии физико-химической эволюции могли возникнуть линейные полимеры, состоящие из почти одинаковых мономерных звеньев. По причинам, изложенным выше, эти линейные полимеры при соответствующих условиях формировали уникальные 3D-структуры. Химическая модификация мономерных звеньев (например, боковых групп) этих полимеров в условиях давления отбора [8], которое направлено на сохранение способности к формированию уникальных 3D-структур могло привести, конечном счете, к набору полипептидов и иных биополимеров, имеющих функциональное значение. Такой механизм формирования молекулярного кода может проектироваться не только на полипептиды, но и на линейные полимеры другой химической природы, включая нуклеотиды или иные структуры, которые оказались, в конечном счете, способными к образованию простейших репликаторов [9].

On the physical basis of the amino acid code for the 3D structure of proteins

Shaitan K.
M.V. Lomonosov Moscow State University, Moscow, Russia
shaytan49@yandex.ru

Key words: linear polymers; polypeptides; multidimensional energy surface topography; folding dynamics; effects of viscosity; configuration space topology and symmetry with respect to rearrangements of identical monomer units; denaturation temperature

Motivation and Aim: It is established that the spatial structure of proteins is determined by the amino acid sequence and corresponds to the minimum free energy of the macromolecule [1]. There are some general ideas that the stability of the 3D-structure requires the presence of an energy gap between the native conformation and the nearest metastable conformations of the chain [2], as well as some relatively smooth topography of the multidimensional energy funnel (the principle of minimum frustration [3]) to overcome the Levinthal paradox [4]. A description of the current state of the problem can be found in a recent review [5]. However, the specific physical aspects that determine the folding process have long remained unclear due to the complexity of the object and the lack of adequate mathematical tools to describe the folding process. Computer methods for predicting the spatial structure of proteins, which are not based on the physics of folding, but use AI technologies, have gained popularity [6]. These algorithms work with varying success, mainly for homologous structures. We proceed from the presumption of the primacy of the physical picture and patterns principle for biological processes, no matter how complex they may seem at the moment [7]. Discovering the physical mechanisms of the organization of the amino acid code for the 3D-structure of proteins is of fundamental importance for understanding the physicochemical foundations of life and can significantly change the existing research paradigm in structural biology. The ideas being developed relate not only to

polypeptides, but also to the formation of 3D-structures of other linear polymers and biopolymers.

Methods and Algorithms: The concepts being developed are based on the structure of linear polymers, the conformation of which is determined by rotations along dihedral angles around chemical bonds. The contribution of vibrations in the valence degrees of freedom to the change in configurations of macromolecules is negligible and is not considered. The topology of the configuration space of a linear polymer is a multidimensional torus. The potential energy (PE) function is defined on a multidimensional torus and is represented as an expansion into a multidimensional Fourier series. The presence of identical or almost identical monomer units in a long polymer chain leads to symmetry of the PE surface with respect to the permutation of the corresponding angular variables. This leads to the dependence of the expansion coefficients on the invariants of algebraic vectors of harmonic numbers, which makes it possible to model specific topographies of multidimensional energy surfaces and then the kinetic and thermodynamic effects. The movement of a representative point (RP) along the multidimensional surface of the PE is described by the equations of mechanics in a viscous medium. In liquid media (with a viscosity on the order of that of water), the inertial terms for atomic motions can be neglected. Using the methods of multidimensional geometry and taking into account the asymptotic properties of functions defined on multidimensional spheres, it is possible to establish a number of general patterns for the rules for RP motion on the multidimensional PE surface. The resulting rules for the RP motion are confirmed by the molecular dynamic simulations.

Results: We can formulate the following rules for the RP movement along the multidimensional energy surface in a viscous medium, which based on kinematic connections for linear polymer nods during the folding [7].

1. The rate of decrease in PE on the mechanical trajectory is maximum compared to other virtual movements.
2. The rate of energy dissipation on the mechanical trajectory is minimal compared to other virtual movements. There is no contradiction here with point 1, since statements 1 and 2 operate in different regions of the phase space of the macromolecule.
3. The rate of energy dissipation, as well as the rate of PE decrease are almost uniformly "spread" over the macromolecule degrees of freedom. This leads to smooth trajectories for the RP motion with the avoidance of sharp changes in the PE relief over a relatively small number degrees of freedom (the "beginner skier" principle [7]).
4. As a consequence, there is a correlation between conformational movements in polymer chain and to form helical structures tendency.

The multidimensional PE surface topography for a linear polymer chain with identical or almost identical monomer units, the following patterns arise [7].

1. The PE surface for linear polymers (built from almost identical units) can be described in terms of two generalized variables: the sum of squared deviations of torsion angles from the PE global minimum point and the sum of the deviations of angles from the PE global minimum point.
2. The topography of the PE surface is a system of nested energy funnels, separated by energy barriers, the height of which decreases with distance from the global minimum point. The central funnel is the deepest.
3. For given values of the generalized variables, the entropy and free energy of the polymer chain are calculated. In this case, a characteristic temperature parameter T_0 arises, which corresponds to the specific negative gain in potential energy during chain

folding per 1 conformational degree of freedom in the monomer unit. At $T>0.26\ T_0$, the folded state becomes unstable. If the denaturation temperature is about 60^0C as in the situation of alpha amino acids (2 degrees of freedom per monomer in the main chain), the negative gain in potential energy per 1 monomer should be approximately 5 kcal/mol, which is close to the hydrogen bond energy.

Conclusion: The above-mentioned regularities of the topography of the PE surface and the rules of the RP movement provide a logical explanation for many heterogeneous experimental data for the kinetics and thermodynamics of the folding and refolding processes [7]. As for the physical basis for the emergence of the amino acid code for the 3D structure of proteins, we formulate the following hypothesis. At the early, prebiological stage linear polymers consisting of almost identical monomer units could arise. For the reasons outlined above, these linear polymers formed unique 3D structures under appropriate conditions. Chemical modification of monomer units (e.g. side groups) of these polymers under selection pressure [8], which is aimed as preserving the ability to form unique 3D structures, could ultimately lead to a set of polypeptides and other biopolymers with functional significance. Such a mechanism for the formation of a molecular code can be designed not only for polypeptides, but also for linear polymers of a different chemical nature, including nucleotides or other structures that ultimately turned out to be capable of forming the simplest replicators [9].

Список литературы/References

1. Anfinsen C.B. Principles that Govern the Folding of Protein Chains. *Science*. 1973;181(4096):223-230. doi 10.1016/S0021-9258(18)64176-6
2. Šali A., Shakhnovich E., Karplus M. How does a protein fold? *Nature*. 1994;369(6477):248-251. doi 10.1038/369248a0
3. Onuchic J.N., Wolynes P.G. Theory of protein folding. *Curr Opin Struct Biol.* 2004;14(1):70-75. doi 10.1016/j.sbi.2004.01.009
4. Levinthal C. Are there pathways for protein folding? *J Chim Physique.* 1968;65(1):44-45. doi 10.1051/jcp/1968650044
5. Finkelstein A.V., Bogatyreva N.S., Ivankov D.N., Garbuzynskiy S.O. Protein folding problem: Enigma, paradox, solution. *Biophys Rev*. 2022;14(6):1255-1272. doi 10.1007/S12551-022-01000-1
6. Callaway E. What's next for the AI protein-folding revolution. *Nature.* 2022;604:234-239. doi 10.1038/d41586-022-00997-5
7. Шайтан К.В. Почему белок сворачивается в уникальную 3D-структуру? И не только это... *Химическая физика*. 2023;42(6):40-62. doi 10.31857/S0207401X23060109
   [Shaitan K.V. Why Do Proteins Fold into Unique 3D-Structures? And Other Questions... *Russ J Phys Chem. B*. 2023;17(3):550-570. doi 10.1134/S1990793123030259 (in English)]
8. Wong M.L. et al. On the roles of function and selection in evolving systems. *PNAS*. 2023;120(43):e2310223120. doi 10.1073/pnas.2310223120
9. Szostak J. How Did Life Begin? *Nature*. 2018;557:S13-S15. doi 10.1038/d41586-018-05098-w

doi 10.18699/bgrs2024-3.1-28

Структура и свойства тиоцианатдегидрогеназ из различных источников

Шипков Н.С.*, Тихонова Т.В., Дергоусова Н.И., Попов В.О.
Федеральный исследовательский центр «Фундаментальные основы биотехнологии» РАН, Москва, Россия
** shipkov1995@yandex.ru*

Ключевые слова: тиоцианат; тиоцианатдегидрогеназа; сероокисляющие бактерии; цианатный путь; медь-содержащие ферменты

Мотивация и цель: Тиоцианат является токсичным соединением, образующимся в как в природных, так и в промышленных процессах. Химическая стабильность и токсичность данного соединения обусловливают необходимость поиска способов очистки промышленных сточных вод от тиоцианата. Перспективным направлением представляется биоремедиация сточных вод с помощью тиоцианат-утилизирующих организмов.
Тиоцианатдегидрогеназа – медь-содержащий фермент, катализирующий окисление тиоцианата с образованием цианата, серы и с переносом двух электронов на внешний акцептор. Первая тиоцианатдегидрогеназа (TcDH) была выделена из галоалкалофильной сероокисляющей бактерии *Thioalkalivibrio paradoxus*. Активный центр TcDH из *Tv. paradoxus* (tpTcDH) содержит набор из 10 консервативных аминокислотных остатков, обеспечивающих поддержание структуры трехъядерного медного кластера и каталитические свойства tpTcDH [1]. Было высказано предположение, что характерный шаблон из 10 каталитически важных аминокислот является маркером принадлежности белка к семейству тиоцианатдегидрогеназ [2]. Целью данной работы было исследование свойств и структуры гомологов tpTcDH с различной степенью гомологии из разных организмов.
Методы и алгоритмы: В работе использовались методы молекулярной биологии для получения рекомбинантных белков, а также различные биохимические методы для характеристики свойств белков.
Результаты Гомологи tpTcDH были обнаружены у более чем 50 микроорганизмов. На основании степени гомологии и присутствия замен в аминокислотном шаблоне гомологи tpTcDH были разделены на три группы. Первая группа – гомологи tpTcDH с идентичностью по аминокислотной последовательности >50 % и без замен в аминокислотном шаблоне, вторая группа – гомологи с идентичностью 49–35 %, но с заменой одной аминокислоты в шаблоне, третья группа – наиболее удаленные гомологи tpTcDH (идентичность <33 %), но без замен в аминокислотном шаблоне. Целью работы являлись получение и характеристика гомологов tpTcDH из каждой группы, проверка важности аминокислотного шаблона для причисления гомологов tpTcDH к семейству тиоцианатдегидрогеназ. Для этого были выбраны TcDH из бактерий *Pelomicrobium methylotrophicum* (pmTcDH) и *Hydrogenobacter thermophilus* (htTcDH) (первая группа), *Tautonia sociabilis* (tsTcDH) (вторая группа) и

Thiohalobacter thiocyanaticus HRh1(ttTcDH) (третья группа). Для всех трех групп бактерий проанализированы генные кластеры, содержащие гены гомологов TcDH. Показано, что бактерия *P. methylotrophicum* способна к росту на тиоцианате как источнике азота, при этом регистрируется экспрессия pmTcDH и фермента цианазы, который разлагает образующийся под действием TcDH цианат на CO_2 и аммоний, использующийся в процессах биосинтеза.

Гомологи tpTcDH из первой группы (pmTcDH и htTcDH) были получены с помощью гетерологичной экспрессии в клетках *E. coli*. Оба белка являются медь-зависимыми ферментами, катализирующими окисление тиоцианата по той же схеме, что и tpTcDH, т. е. являются тиоцианатдегидрогеназами. Охарактеризованы биохимические свойства обоих белков.

Вторая группа гомологов tpTcDH содержит одну замену в аминокислотном шаблоне, остаток H482, который в tpTcDH отвечает за координацию иона Cu^3, заменен на глутамин. Полученная методом гетерологичной экспрессии tsTcDH из *T. sociabilis*, относящаяся ко второй группе, не обладала тиоцианатдегидрогеназной активностью. Способность к росту на тиоцианате была проверена также для бактерии *Methylocapsa acidiphila* B2, содержащей в геноме ген гомолога TcDH из второй группы. Бактерия *M. acidiphila* B2 оказалась не способна к росту на тиоцианате. Таким образом, замены в аминокислотном шаблоне критичны, и белки, содержащие такую замену, не являются тиоцианатдегидрогеназами.

Для исследования третьей группы гомологов tpTcDH была выбрана бактерия *T. thiocyanaticus*, для которой ранее была показана способность расти на тиоцианате как единственном источнике энергии и азота. Фермент ttTcDH (идентичность по аминокислотной последовательности с tpTcDH 32 %), выделенный из периплазматической фракции *T. thiocyanaticus*, выращенной на минеральной среде в присутствии тиоцианата, представлял собой смесь свободного фермента и прочного комплекса ttTcDH с тиоредоксин-подобным белком (TLP), ген которого предшествует гену ttTcDH. Препарат содержал до 2.9 иона меди в пересчете на TcDH. Тиоцианатдегидрогеназной активностью обладал только комплекс ttTcDH-TLP, свободный ttTcDH не проявлял каталитических свойств.

Для дальнейших исследований были получены рекомбинантные препараты ttTcDH и TLP. Поскольку рекомбинантный ttTcDH был нестабилен и склонен к агрегации, были проведены широкий скрининг и оптимизация методов получения и стабилизации целевого белка, позволившие увеличить выход ttTcDH и избавиться от высокомолекулярных агрегатов. Свободный рекомбинантный ttTcDH, как и нативный, не обладал тиоцианатдегидрогеназной активностью после насыщения ионами меди. Каталитическая активность ttTcDH появляется только при насыщении ионами меди в присутствии TLP. При этом образуется комплекс ttTcDH-TLP, который наблюдается на гель-фильтрации. Обнаружена способность TLP связывать ионы меди. Возможно, TLP участвует во встраивании ионов меди в активный центр ttTcDH, это подтверждается тем, что насыщение ионами меди ttTcDH в отсутствие TLP приводит к встраиванию только одного из трех ионов меди.

Анализ ближайшего окружения гена TcDH у микроорганизмов разных классов показал, что у тех бактерий, где гомологи TcDH имеют консервативный шаблон из 10 аминокислотных остатков (первая и третья группы гомологов), рядом с геном TcDH расположены гены медь-связывающих белков, ответственных за

перенос и встраивание ионов меди в активный центр с образованием каталитически компетентной формы фермента. В случае организмов второй группы, где у предполагаемой TcDH присутствует замена одного из остатков аминокислотного шаблона, гены белков – переносчиков ионов меди – в ближайшем окружении гена TcDH отсутствуют. Поскольку из наших данных следует, что такая форма TcDH неактивна, то клетке нет необходимости создавать систему транспорта ионов меди к заведомо дефектному белку.

Выводы:

1. Получены и охарактеризованы гомологи tpTcDH из четырех организмов.
2. Подтверждено, что шаблон из 10 аминокислотных остатков, формирующих активный центр tpTcDH, отвечает за принадлежность гомологов tpTcDH к тиоцианатдегидрогеназам.
3. Все изученные тиоцианатдегидрогеназы требуют присутствия трех ионов меди в активном центре для каталитической активности.
4. Нативный ttTcDH находится в двух формах: свободной и в прочном комплексе с тиоредоксин-подобным белком (TLP). Только комплекс ttTcDH-TLP обладает ферментативной активностью. Свободная ttTcDH неактивна и нестабильна.
5. Оптимизирован способ получения и стабилизации рекомбинантной ttTcDH, позволивший получить гомогенный и монодисперсный препарат белка.
6. Насыщение ttTcDH ионами меди приводит к встраиванию только одного из трех ионов, и фермент оказывается неактивен. Насыщение ttTcDH ионами меди в присутствии TLP приводит образованию каталитически активной формы фермента.

Structure and properties of thiocyanate dehydrogenase from various sources

Shipkov N.S.*, Tikhonova T.V., Dergousova N.I., Popov V.O.

Federal Research Centre "Fundamentals of Biotechnology", RAS, Moscow, Russia

** shipkov1995@yandex.ru*

Key words: thiocyanate; thiocyanate dehydrogenase; sulfur-oxidizing bacteria; cyanate pathway; copper-containing enzymes

Motivation and Aim: Thiocyanate is a toxic compound formed in both natural and industrial processes. The chemical stability and toxicity of thiocyanate necessitate the search for ways to purify industrial wastewater from thiocyanate. Bioremediation of wastewater using thiocyanate-recycling organisms seems to be a promising direction.

Thiocyanate dehydrogenase is a copper-containing enzyme that catalyzes the oxidation of thiocyanate to form cyanate, sulfur and transfer two electrons to an external acceptor. The first thiocyanate dehydrogenase (TcDH) was isolated from the haloalkalophilic sulfur-oxidizing bacterium *Thioalkalivibrio paradoxus*. The active center of TcDH from *Tv. paradoxus* (tpTcDH) contains a set of 10 conserved amino acid residues that maintain the structure of a three-core copper cluster and the catalytic properties of tpTcDH [1]. It has been suggested that the characteristic pattern of 10 catalytically important amino acids is a marker of a protein belonging to the thiocyanate

dehydrogenase family [2]. The purpose of this work was to study tpTcDH homologues from different organisms.

Methods and Algorithms: The work used methods of molecular biology to obtain recombinant proteins, and various biochemical methods to determine the properties of proteins.

Results: tpTcDH homologues have been found in more than 50 microorganisms. Based on the degree of homology and the presence of substitutions in the amino acid template, tpTcDH homologues were divided into 3 groups. The first group contains tpTcDH homologues with homology >50 % with no substitutions in the amino acid template, the second group includes homologues with 49–35 % homology with one amino acid substitution in the amino acid template, and the third group involves the most distant tpTcDH homologues (homology <33 %), but without substitutions in the amino acid template. The aim of the work was to obtain and characterize tpTcDH homologues from each group, to verify the importance of the amino acid template for classifying tpTcDH homologues to the thiocyanate dehydrogenase family. For this purpose TcDH homologues from bacteria *Pelomicrobium methylotrophicum* (pmTcDH) and *Hydrogenobacter thermophilus* (htTcDH) (first group), *Tautonia sociabilis* (tsTcDH) (second group) and *Thiohalobacter thiocyanaticus* HRh1(ttTcDH) (third group) were selected. Gene clusters containing genes of the TcDH homologue were analyzed for all three groups of bacteria.

The work shows that the bacterium *P. methylotrophicum* is capable of growing on thiocyanate as a nitrogen source, while the expression of pmTcDH and the cyanase enzyme is recorded, which decomposes cyanate formed under the action of TcDH into CO_2 and ammonium used in biosynthesis processes.

tpTcDH homologues from the first group (pmTcDH and htTcDH) were obtained by heterologous expression in *E. coli* cells. Both proteins are copper-dependent enzymes that catalyze the oxidation of thiocyanate according to the same scheme as tpTcDH, that is, they are thiocyanate dehydrogenases. The biochemical properties of both proteins have been studied.

The second group of tpTcDH homologues contains one substitution in the amino acid template, the H482 residue, which in tpTcDH is responsible for coordinating the Cu^3 ion, is replaced by glutamine. The tsTcDH from *T. sociabilis* obtained by heterologous expression, belonging to the second group, did not demonstrate thiocyanate dehydrogenase activity. The ability to grow on thiocyanate was also tested for the bacterium *Methylocapsa acidiphila* B2, which contains the gene of TcDH homologue from the second group in the genome. The bacterium *M. acidiphila* B2 was unable to grow on thiocyanate. Thus, substitutions in the amino acid template are critical, and proteins containing such a substitution are not thiocyanate dehydrogenases.

To study the third group of tpTcDH homologues, the bacterium *T. thiocyanaticus* was selected, for which the ability to grow on thiocyanate as the only source of energy and nitrogen was previously shown. The ttTcDH enzyme (amino acid sequence identity with tpTcDH 32 %) isolated from the periplasmic fraction of *T. thiocyanaticus* grown on a mineral medium in the presence of thiocyanate was a mixture of a free enzyme and a strong complex of ttTcDH with a thioredoxin-like protein (TLP), the gene of which precedes the ttTcDH gene. The preparation contained up to 2.9 copper ions in terms of TcDH. Thiocyanate dehydrogenase activity was possessed only by the ttTcDH-TLP complex, free ttTcDH did not exhibit catalytic properties.

Recombinant drugs ttTcDH and TLP were obtained for further studies. Since recombinant ttTcDH was unstable and prone to aggregation, extensive screening and optimization of methods for obtaining and stabilizing the target protein were carried out, which made it possible to increase the yield of ttTcDH and get rid of high-molecular aggregates. Free recombinant ttTcDH, like the native one, did not have thiocyanate dehydrogenase activity after saturation with copper ions. The catalytic activity of ttTcDH appears only when saturated with copper ions in the presence of TLP. In this case, the ttTcDH-TLP complex is formed, which is observed on gel filtration. The ability of TLP to bind copper ions has been discovered. It is possible that TLP is involved in the incorporation of copper ions into the active center of ttTcDH, this is confirmed by the fact that saturation with copper ions of ttTcDH in the absence of TLP leads to the incorporation of only 1 out of 3 copper ions.

Analysis of the immediate environment of the TcDH gene in microorganisms of different classes showed that in those bacteria where TcDH homologues have a conservative template of 10 amino acid residues (the first and third groups of homologues), next to the TcDH gene there are genes of copper-binding proteins responsible for the transfer and incorporation of copper ions into the active center to form a catalytically competent form the enzyme. In the case of organisms of the second group, where the proposed TcDH has a replacement of one of the residues of the amino acid template, the genes of copper ion carrier proteins in the immediate environment of the TcDH gene are absent. Since our data show that this form of TcDH is inactive, the cell does not need to create a system for transporting copper ions to a known defective protein.

Conclusion:

1. tpTcDH homologues from 4 organisms were obtained and characterized.
2. It has been confirmed that a template of 10 amino acid residues forming the active center of tpTcDH is responsible for the affiliation of tpTcDH homologues to thiocyanate dehydrogenases.
3. All studied thiocyanate dehydrogenases require the presence of 3 copper ions in the active center for catalytic activity.
4. Native ttTcDH is found in two forms: free and in a strong complex with a thioredoxin-like protein (TLP). Only the ttTcDH-TLP complex has enzymatic activity. Free ttTcDH is inactive and unstable.
6. The method of obtaining and stabilizing recombinant ttTcDH has been optimized, which made it possible to obtain a homogeneous and monodisperse protein preparation
7. Saturation of tch with copper ions leads to the incorporation of only one of the three ions, and the enzyme is inactive. Saturation of ttTcDH with copper ions in the presence of TLP leads to the formation of a catalytically active form of the enzyme.

Список литературы/References

1. Tikhonova T.V. et al. Trinuclear copper biocatalytic center forms an active site of thiocyanate dehydrogenase. *Proc Natl Acad Sci USA*. 2020;117(10):5280-5290
2. Tsallagov S.I. et al. Comparative Genomics of Thiohalobacter thiocyanaticus HRh1T and Guyparkeria sp. SCN-R1, Halophilic Chemolithoautotrophic Sulfur-Oxidizing Gammaproteobacteria Capable of Using Thiocyanate as Energy Source. *Front Microbiol*. 2019;10:898

doi 10.18699/bgrs2024-3.1-29

Implementation of Steered Molecular Dynamics (SteeredMD) in the Schrodinger molecular dynamics software

Bashilov A.[1, 2]*, Gavryushov S.[2]

[1] *Skolkovo Institute of Science and Technology, Moscow, Russia*

[2] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

* *anton_bashilov@mail.ru*

Key words: Steered Molecular Dynamics; GPCR; Desmond; CB1

Motivation and Aim: The goal of this study was to develop a program called GuidedMD, which implements Steered Molecular Dynamics (SteeredMD) for the Schrödinger suite of programs (Schrödinger Inc., New York, NY, USA). This tool can be applied to explore conformational changes in proteins and protein-ligand interactions. The method was essential for studying allosteric modulation in the cannabinoid type 1 (CB1) G-protein-coupled receptor (GPCR) [1]. The standard biasing force plugin does not support the use of complex trajectories of atom displacements chosen for SteeredMD, therefore the main requirements for GuidedMD were flexibility (to perform simulations along complex trajectories and easily select steered atoms) and the capability to evaluate free energy changes.

Methods and Algorithms: GuidedMD utilizes Desmond [2] to conduct SteeredMD simulations. The biasing potential is implemented as a time-dependent moving harmonic spring. With this approach, an external steering force is applied to a set of selected atoms (F). A second selected set of atoms (Basis B) is used to restrain and retain the whole molecular system. The main routine involves a cyclic call of Desmond MD for each movement of the biasing potential (after each MD simulation output files are processed with a script to modify the coordinates of the restraining potentials for selection F). These modified files are then used as input for the subsequent iteration. The mean force of the system and the transformation of the system from state 1 to state 2 by pulling N atoms can be estimated as follows:

$$\langle \vec{F_i} \rangle = k \langle \Delta \vec{r_i} \rangle$$

$$\Delta G_{2,1} = -\sum_{i=1}^{N} \int_1^2 \langle \vec{F_i} \rangle d\vec{r}$$

To perform a SteeredMD simulation using GuidedMD, one must first prepare the input files for Desmond (e. g., via System Builder followed by Maestro's GUI, specifying the restrained atoms for both F and B selections). Secondary, trajectories for the restraining potential offsets should be generated (manually, or using a part of the GuidedMD TaskGen scripts). Thirdly, main routine script (CmsR) should be configured according to the short manual and executed. After simulation a BuildPlot script (part of GuidedMD) can be executed to get the results of SteeredMD: a PMF plot for all steered atoms and ΔG for the SteeredMD run. GuidedMD has been validated by several tests. In particular, a potential of mean force (PMF) for Cl^--Na^+ pair in vacuum was calculated. Chlorine ion (B) was restrained with the spring constant of 500 kcal/mol/Å^2, the moved

sodium ion (F) was restrained by 50 kcal/mol/Å^2. The system was built and equilibrated according to the protocol for CB1 receptor [1]. Number of MD steps (I) was 180, each MD run was for 10 ns at 300K in NVT ensemble (Nose-Hoover chain thermostat, relaxation 1 ps). For each iteration (i) coordinate of restraining potential ($\vec{r}$) for the sodium ion was set using equation:

$$(\vec{r}_{Na^+})_{new} = (\vec{r}_{Cl^-} - \vec{r}_{Na^+})/(I - i)$$

Also, the PMF for Na^+-Cl^- pair was conducted in SPC/E water. The protocol was similar with some modifications described below. Chlorine and sodium ions were placed in a box (70Å × 20Å × 20Å) so that they were equidistant from the box boundaries, and the distance between them was 20 Å in the x direction. Additionally, three oxygen atoms from the SPC/E water were restrained (for box alignment). The ensemble was NPT: using the Martyna-Tobias-Klein barostat, with a relaxation time of 2 ps and a pressure of 1.01325 MPa. The number of MD steps was 180.

Results: The accuracy of the test PMF computation in vacuum (in comparison with theoretical curve) was satisfactory: the relative accuracy was less than 0.1 %. As for simulation in SPC/E water, the shape of the computed curve is similar to published before [3]. The computed dielectric constant (Fig. 1) for SPC/E water (69) is close to published one (71.8) [4].

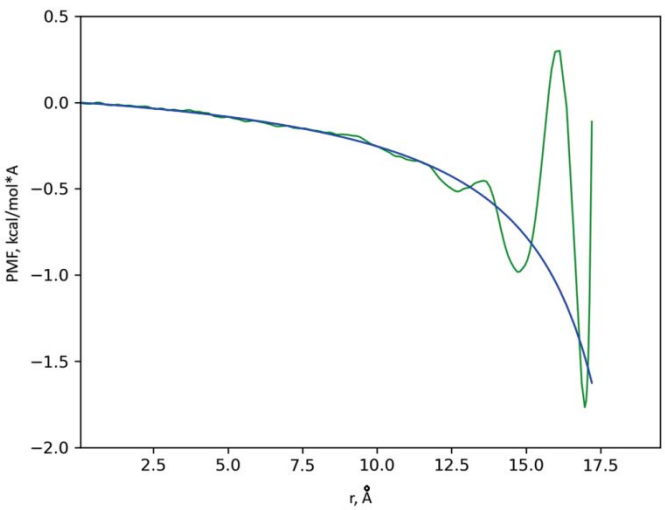


Fig. 1. PMF curve for sodium and chlorine ions in SPC/E water. Blue is theoretical PMF curve in vacuum normalized on $\varepsilon_r = 69$, green is result of computation via GuidedMD

An example of guided structure modification was previously described in detail by Gavryushov et al. [1]. Figure 2 describes the vectors of steering displacements at the structural transformation of the membrane domain of CB1 receptor from its inactive conformation to the active state. The steering forces are applied to a part of the transmembrane helix (TM7), with atoms shown as van der Waals (vdW) spheres, allowing the transformation of CB1 from the inactive to the active state *'in silico'*. The SteeredMD transformation of the CB1 domain immersed into membrane included 20 iterations, each for 10 ns. The energy of transformation in the presence of an agonist

ligand was approximately 0 kcal/mol, while without the ligand, it was above 20 kcal/mol. The latter value could estimate the transformation free energy difference.

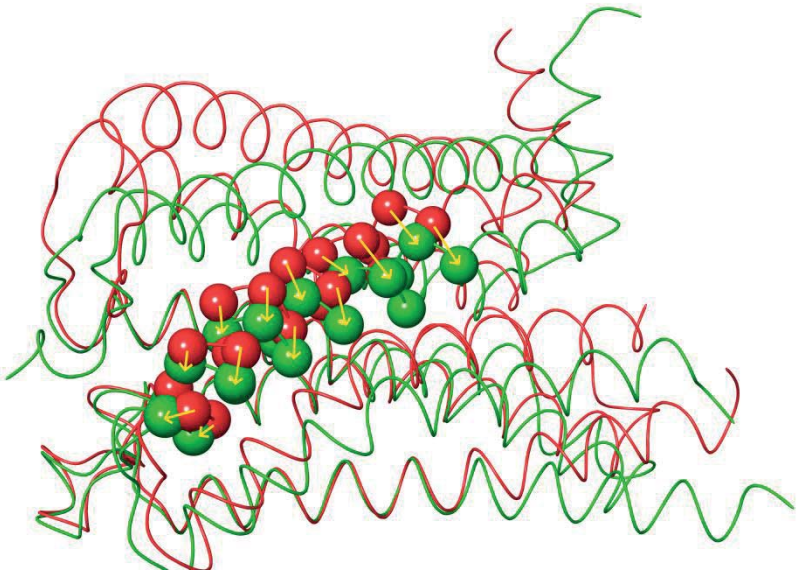


Fig. 2. Steering vectors for CB1 transformation. Red – inactive CB1 (from PDB code 5TGZ), green – active CB1 (relaxed structure from PDB code 5XRA). Yellow steering vectors show displacements of the alpha carbon atoms moved by the steering external forces

Conclusion: A GuidedMD is a flexible and user-friendly tool for performing SteeredMD using the Desmond software. With this developed software, it is possible to accomplish such tasks as: PMF calculations, guided structure transformations, rebuilding of missing chains in X-ray structures of proteins, and relative free energy calculations (using the equation above as well as the Jarzynski equation [5]).

Funding: The study had no funding.

References

1. Gavryushov S., Bashilov A., Cherashev-Tumanov K.V., Kuzmich N.N., Burykina T.I., Izotov B.N. Interaction of Synthetic Cannabinoid Receptor Agonists with Cannabinoid Receptor I: Insights into Activation Molecular Mechanism. *Int J Mol Sci.* 2023;24(19):14874. doi 10.3390/ijms241914874
2. Schrödinger Release 2019-2: Desmond Molecular Dynamics System. New York: D.E. Shaw Research, 2019
3. Gavryushov S., Per L. Effective interaction potentials for alkali and alkaline earth metal ions in SPC/E water and prediction of mean ion activity coefficients. *J Phys Chem B.* 2006;110(22):10878-10887. doi 10.1021/jp056871i
4. Reddy M., Rami Berkowitz M. The dielectric constant of SPC/E water. *Chem Phys Lett.* 1989;155(2):173-176. doi 10.1016/0009-2614(89)85344-8
5. Jarzynski C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physi Rev E.* 1997;56(5):5018. doi 10.1103/PhysRevE.56.5018

doi 10.18699/bgrs2024-3.1-30

Exploring human mitochondrial inorganic pyrophosphatase functionality: insights from molecular dynamics and effects of the deleterious Met94Val mutation

Bezpalaya E.[1]*, Rodina E.[1], Vorobyova N.[2], Kurilova S.[2]

[1] *Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia*

[2] *A.N. Belozersky Institute of Physico-Chemical Biology, Moscow, Russia*

* *bezpalaya.katya@gmail.com*

Key words: inorganic pyrophosphatase; mitochondrial; mutation; molecular dynamics

Motivation and Aim: Mitochondrial inorganic pyrophosphatase (PPA2), encoded by the *ppa2* gene, is an essential yet poorly characterized enzyme [1]. Naturally occurring mutations in its human gene disrupt the regulation of mitochondrial membrane potential, leading to mitochondrial dysfunction and resulting in severe heart conditions [2, 3]. The main focus of our study revolves around two human PPA2 (hPPA2) proteins: the wild type (WT) and one of the pathogenic variants, Met94Val. To broaden our understanding during the research, we also incorporated two additional subjects, analogous PPA2 from yeast *O. parapolymorpha* (WT OpPPA2 and Met52Val OpPPA2). Our aim is to elucidate the effects of the mutation on the proteins' behavior and identify patterns of its conformational changes using molecular dynamics (MD).

Methods and Algorithms: For human proteins, we modelled structures using I-TASSER. Two cofactor Mg^{2+} ions were modelled using a crystal structure of homologous PPase. Subunit structure was duplicated to create a dimer. For *O. parapolymorpha* proteins, we used crystal structure of dimeric OpPPA2 obtained earlier. Mutations were introduced by altering the residues using UCSF Chimera. Molecular dynamics (MD) simulations of 100 ns duration were conducted using the AmberTools22 package. Each structure was embedded in a truncated octahedron unit containing 55.5 M water molecules and 0.15 M salt concentration. An energy cut-off parameter of 10 Å was set, with a buffer distance of 12 Å. After minimization the system underwent a heating process from 0 to 300 K. Production simulations were performed under constant pressure at 300 K.

Results: Although the methionine residue is conserved among homologs, it neither holds a known functional role nor directly relates to the active site. This residue is a center of PPA2 hydrophobic core. Methionine in such positions is responsible for sustaining protein stability and providing flexibility because of its long, sulfur-containing side-chain. To assess the proteins stability, we obtained RMSD values for all 4 objects backbone atoms. Throughout the simulation, RMSD consistently remained below ~3 Å. Additionally, RMSD of mutants and wild-type proteins showed similar levels, indicating comparable global system stability. To evaluate residue flexibility, we calculated backbone atoms RMSF. Both OpPPA2 and hPPA2 enzymes exhibited similar RMSF patterns across protein domains. As expected, a significant decrease in flexibility occurred at the hydrophobic beta-barrel region, where Met94 of hPPA2 (Met52 of OpPPA2) is located. Interestingly, the mean RMSF of the Met52Val OpPPA2 dropped by 2.1 Å compared to the wild type, while the RMSF of the Met94Val hPPA2 was only slightly lower. We supposed that Val52 in the OpPPA2 mutant might not offer enough

flexibility due to its branched aliphatic chain. Further research is needed to understand why the hPPA2 experiences fewer effects, despite structural alignment showing no spatial differences in hydrophobic cores between OpPPA2 and hPPA2. Most active site residues showed decreased RMSF values upon mutation in both proteins, with the difference being more significant in the OpPPA2. To check whether any atomic positions rearrangement happens in the active site of the mutants compared to WT, we aligned the respective structures of the representative MD frames. We found little difference, except for Glu97 in hPPA2 and analogous Glu55 in OpPPA1. As these residues are involved in Mg^{2+} binding, their displacement may impact all catalytic stages. Using the principal component method, we analyzed protein conformations reached during the 100 ns simulation. Projection of phase space on the first two components indicated two main conformations clusters for human enzymes and transitions between them. Similarly, for the Met52Val there are two clusters far from each other. However, WT clusters are close to each other. To evaluate the scope of atoms movement of each protein, we acquired the eigenvalues of the covariance matrices which correspond to the first principal component. While eigenvalues were comparable for human proteins, for the WT and Met52Val OpPPA2 their ratio was ~1:15 respectively which indicated a significant difference in conformational flexibility. Visualizing PCA modes trajectories revealed that the most pronounced mode is subunits' relative movement. Analyzing contacts of methionine with surrounding residues revealed minimal change for OpPPA2. Here and onwards, by residue contacts, we mean the presence of their atoms within a sphere with a radius of 7 Å, as within this distance, electrostatic and van der Waals interactions, as well as contacts driven by hydrophobic effects, may occur. However, the human Met94Val variant showed a threefold reduction in contacts with crucial Tyr142 compared to WT. The explanation may be that, while directed to the tyrosine 142 backbone atoms, methionine acts as its anchor, but valine cannot act the same because of its short chain. Tyr142 plays role in substrate and product binding and is a second cofactor ligand, therefore, mutation may lead to incorrect positioning of the residue within the active site. Moreover, number of contacts with the above-mentioned Glu97 decreases a lot, suggesting that its anchoring may be affected too.

Conclusion: RMSD and PCA results obtained from MD trajectories suggest that the global dynamics of hPPA2 does not significantly change upon Met94Val mutation. On the other hand, analogous mutation Met52Val significantly decreases residues' flexibility in *O. parapolymorpha* PPA2. Replacement of long sulfur-containing methionine chain by short valine in hPPA2 affects its interaction with the backbone of essential residues Tyr142 and Glu97, which may lead to the disruption of active site preorganization and catalysis.

Funding: The study is supported the Russian Scientific Foundation Grant (No. 23-24-00177).

References

1. Lundin M., Baltscheffsky H., Ronne H. Yeast PPA2 gene encodes a mitochondrial inorganic pyrophosphatase that is essential for mitochondrial function. *J Biol Chem.* 1991;266:12168-12172. doi 10.1016/S0021-9258(18)98875-7
2. Genthe W., Donnelly C., Ezon D., Fettig V., Ganesh J., Marin-Valecia I., Gelb B. PPA2 Deficiency in 2 Sisters. *JACC Case Rep.* 2023;24:1024. doi 10.1016/j.jaccas.2023.102024
3. Phoon C.K.L., Halvorsen M., Goldstein D.B., Rabin R., Cecchin F., Crandall L., Devinsky O. Sudden unexpected death in asymptomatic infants due to PPA2 variants. *Mol Genet Genomic Med*. 2020;8(1):e1008. doi 10.1002/mgg3.1008

doi 10.18699/bgrs2024-3.1-31

Calculating of SARS-CoV-2 RBD and antibody complex energy by metadynamics in MARTINI 3 for evaluating complex free energy

Chuiko Ya.V.[1]*, Golovin A.V.[2]

[1] *Sirius University of Science and Technology, Sochi, Russia*

[2] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

* *chujko.yav@learn.siriusuniversity.ru*

Key words: antibody design; generative models; molecular dynamics; metadynamics

Motivation and Aim: Neutralizing monoclonal antibodies (MAbs) have been found to be effective therapeutic agents in the treatment of a host of viral and autoimmune diseases. This has been particularly highlighted since the outbreak of the 2019 coronavirus infection (COVID-19), which is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In response to this global health crisis, the Food and Drug Administration (FDA) has given its approval to several drugs that are based on MAbs [1]. The emergence of new variants of the virus has presented a significant hurdle in the ongoing fight against COVID-19. Some of these variants have shown resistance to the existing neutralizing MAbs, reducing their overall efficacy [2]. There's an urgent need for new strategies for rapid antibody design to respond to emerging variants and maintain our therapeutic effectiveness.

Nowadays, there is a rising trend towards the utilization of diffusion machine learning models in the field of the design of innovative protein sequences [3]. These advanced models allow for the intricate redesigning of the antibody constant determining regions (CDR) loops. This is a significant step in the field as it enables us to specifically tailor these CDR loops to correspond with the antigen epitope we require.

To improve the design results and antibody affinity to antigen, we must have initial stable structures of the antigen antibody complex, in which the antibody will target different epitopes. These states can be obtained by the methods of molecular dynamics. One of them, rapid and effective metadynamics, allows one to quickly and efficiently scan different states of a system by adding Gaussian potentials to local free energy minima during a trajectory [4]. Adding to this, representation of structures in the Martini force field together with the use of metadynamics speeds up FES calculations for protein-protein complexes [5]. Therefore, using the metadynamics method in MARTINI 3 atoms representation can allow us to estimate and range hundreds of results of structures obtained from generative diffusion models.

In this report, we detail the application of metadynamics to the most common antibody and SARS-CoV-2 RBD in the Martini 3 force field. We initially created two complexes where the antibody targets the two most likely neutralizing receptor binding domain epitopes. We then performed metadynamics on each complex and identified the states corresponding to the free energy minima.

Methods and Algorithms: SARS-Cov-2 neutralizing MAbs targeted to SARS-CoV-2 S-protein RBD data were downloaded from The Coronavirus Antibody Database CoV-AbDab [6]. Based on them, two antibody-antigen complexes were constructed, in which

the antibody targets the two most common epitopes. Antibodies were constructed with the most common CDR lengths and amino acid sequences by MODELLER [7]. All-atom structures of constructed complexes were presented in coarse grained resolution structures according to the Martini 3.0.beta.3.2 force field (Martini 3) [8 (12)]. Coarse grained structures of two complexes were solvated in triclinic water boxes with ions and were minimized by the steepest descent algorithm. Metadynamics simulations were run under NPT ensemble by GROMACS software with PLUMED. The distance between the antibody and RBD centers of mass (COM) and the dihedral angle characterizing the rotation of the antibody relative RBD were selected as CV. The final minimum structures complexes were back-mapped to the all atom resolution by the 'backward.py' script [8]. *Results*: After conversion to CG-structures and its preparation metadynamics in two CVs was performed. Reweighted free energy surfaces for epitope-A and epitope-B complexes are shown in Figure 1. First five free energy minima (values less than 0) are pointed out with numbers from 1 to 5. Time frames according to states of complexes for these free energy minima were extracted from the trajectory and clustered.

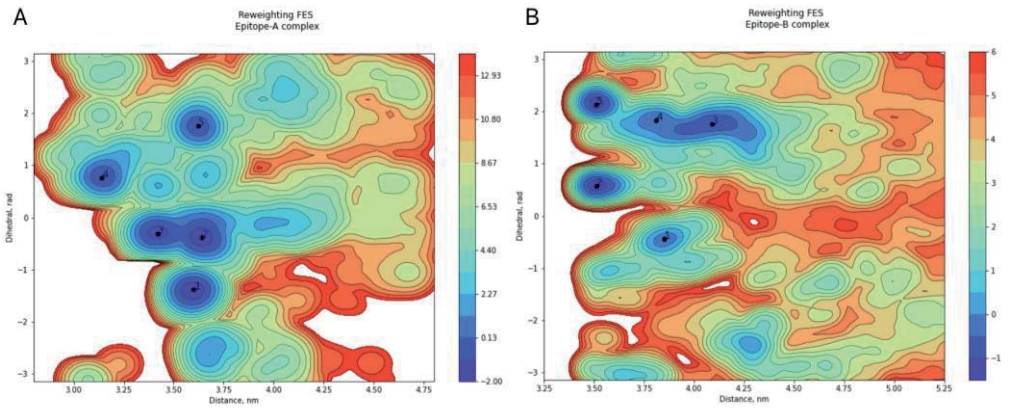


Fig. 1. Reweighting FES plot for epitope-A complex (A) and epitope-B complex (B)

Figure 2 provides an illustrative map detailing the contact residues of the Receptor Binding Domain (RBD) for each minimal state that has been obtained through the process of metadynamics. In the context of the epitope-B complex, the antibody specifically targets the cryptic epitopes site when in minimal states in addition to the main observed epitopes.

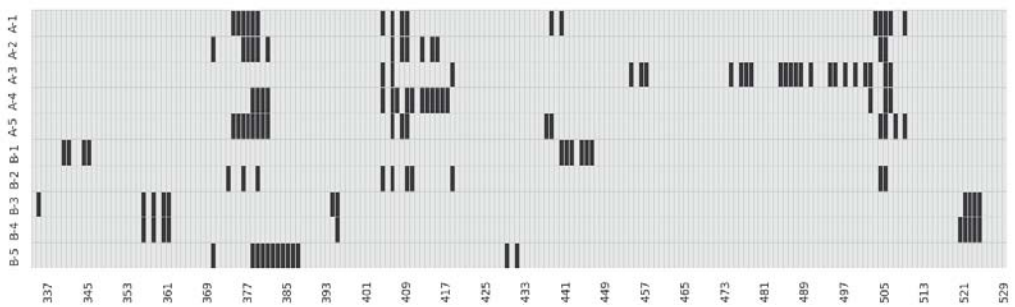


Fig. 2. Representation of RBD residues involved in binding to the antibody for minimal states from epitope-A and epitope-B complexes

Conclusions: According to the values of the collective variables and the obtained conformations of the complexes, the antibody can be targeted to different epitopes during metadynamics, which gives a wide variety of structures on the basis of which new antibodies can be obtained using generative diffusion models. Also, given the high speed of metadynamics in the Martini 3 force field and at the same time effective sampling of the energy space, this approach can be used to evaluate the generated designs.

References

1. GSK sotrovimab letter of authorization 02232022 [Electronic resource]. URL: https://www.fda.gov/media/149532/ (accessed: 15.05.2024)
2. Wang L. et al. Immune evasion of neutralizing antibodies by SARS-CoV-2 Omicron. *Cytokine Growth Factor Rev*. 2023;70:13-25
3. Watson J.L. et al. De novo design of protein structure and function with RFdiffusion. *Nature*. 2023;620(7976):1089-1100
4. Ray D., Parrinello M. Kinetics from Metadynamics: Principles, Applications, and Outlook. *J Chem Theory Comput*. 2023;19(17):5649-5670
5. Lamprakis C. et al. Evaluating the Efficiency of the Martini Force Field to Study Protein Dimerization in Aqueous and Membrane Environments. *J Chem Theory Comput*. 2021;17(5):3088-3102
6. Raybould M.I.J. et al. CoV-AbDab: the coronavirus antibody database. *Bioinformatics*. 2021;37(5):734-735
7. Webb B., Sali A. Protein Structure Modeling with MODELLER. *Methods Mol Biol*. 2017;1654:39-54
8. Wassenaar T.A. et al. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J Chem Theory Comput*. 2014;10(2):676-690

doi 10.18699/bgrs2024-3.1-32

3D modeling of the opsin structure for the gastropod mollusc *Lissachatina fulica* using AlphaFold 2

Dominova I.N.*, Fedotov D.A., Shirina S., Zykin V., Zhukov V.V.
Immanuel Kant Baltic Federal University, Kaliningrad, Russia
** irinadominova@gmail.com*

Key words: Onsin; 3D structure; gastropod mollusk; AlphaFold 2

Motivation and Aim: The study of protein structure and function is fundamental to understanding life processes at the molecular level. In particular, visual pigments, such as opsins, play a key role in the mechanisms of vision in animals [1]. Obtaining accurate models of the spatial structure of opsins makes it possible to reveal the specific features of their functioning. Current knowledge of these proteins' structure has been obtained mainly in studies on vertebrates, arthropods, and cephalopod molluscs. However, there is little data on the structure of visual opsins in gastropod mollusc and none for terrestrial species. Thus, the aim of this work was to model the 3D structure and determine the most likely candidate for the role of the optic opsin in *Lissachatina fulica*.
Methods and Algorithms: In this work, 5 previously predicted *L. fulica* opsins, Afu011750, Afu004575, Afu005002, Afu003765 and Afu022267, were analyzed. The primary sequences of *L. fulica* proteins were taken from Guo's article [2]. AlphaFold 2 algorithm was used to predict 3D structures [3], visualization and analysis of structural homology of the obtained structures with the 3D structure of rhodopsin *Todarodes pacificus* (RCSB Protein Data Bank database) resolved by X-ray structural analysis with an accuracy of 3.70 Å were performed using PyMol. The structural homology analysis took into account the root mean square deviation (RMSD), which measures how much a particular molecular structure deviates from the reference geometry [4].
Results: By analyzing the structural homology and RMSD obtained by comparing the predicted *L. fulica* opsins with *T. pacificus* rhodopsin, the three sequences Afu005002 (RMSD = 0.916 Å), Afu004575 (RMSD = 1.158 Å), and Afu016848 (RMSD = 2.829 Å) were shown to have the lowest RMSD values and thus the highest structural homology. The influence of the nearest amino acid residues on the positions of the main absorption maximum for Afu005002 was also analyzed. As a result, it was found that the greatest contribution to the absorption spectrum and to the spectral shift of the predicted opsin structure of Afu005002 was made by protonation of Glu175.
Conclusion: Thus, we can conclude that Afu005002 is the most likely candidate for the role of a visual opsin in *L. fulica*. Our study opens new opportunities for further research on opsin proteins, including the study of their visual mechanisms.
Funding: The study is supported by grant No. 434-K-23 of the Strategic Academic Leadership Program “Priority 2030” of IKBFU.

References

1. Cronin T.W., Johnsen S., Marshall N.J., Warrant E.J. Visual ecology. Princeton University Press, 2014
2. Guo Y., Zhang Y. et al. N. A chromosomal-level genome assembly for the giant African snail *Achatina fulica*. *GigaScience*. 2019;8(10):giz124. doi 10.1093/gigascience/giz124
3. Mirdita M., Schütze K., Moriwaki Y. et al. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022;19:679-682. doi 10.1038/s41592-022-01488-1
4. Kufareva I., Abagyan R. Methods of protein structure comparison. *Methods Mol Biol.* 2012;857:231-257. doi 10.1007/978-1-61779-588-6_10

doi 10.18699/bgrs2024-3.1-33

“Dynamic molecular portrait” approach in computational structural biology of membrane proteins

Efremov R.G.
Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia
r-efremov@yandex.ru

Key words: protein structure and dynamics; molecular dynamics; protein-membrane interactions; computational molecular modeling

Understanding at the molecular level the structure, dynamics, and mechanisms of functioning of membrane proteins (MPs) is important both from a fundamental and applied point of view – for the development of new drugs and therapies for a number of socially significant diseases. Traditionally, of particular interest in recent years has been the study of receptors and ion channels. In this work, we make focus on two families from these classes of MPs – receptor tyrosine kinases (RTKs) and transient receptor potential (TRP) ion channels. This is due to the presence of experimentally determined models of the spatial structure of a number of these MPs in various functional states. Computer modeling provides important information about the features of MPs, allowing, on the basis of experimentally obtained models, to evaluate the fine details of their conformational dynamics, analyze the evolution of their physicochemical properties, assess the integral membrane effects and the role of individual bound lipids, etc. The results of calculations contribute to the identification at the atomic level of the relationship of these characteristics of MPs of different types with the parameters of their action in the cell. The report presents the results of in silico modeling of a number of proteins from the RTK and TRP family that were taken in several functional states and embedded in hydrated lipid bilayers mimicking the cell membrane. Calculations of molecular dynamics in the all-atom presentation, as well as original computer technologies for mapping and visualization of physicochemical properties of protein-membrane systems were used as the main methods of computational experiment. Special attention was paid to application of our original “dynamic molecular portrait” technology, which permits detailed assessment and pictorial visualization of the combination of structural, dynamic, hydrophobic, electric, and other properties of MPs calculated with high spatial and temporal resolution. The results obtained significantly supplement the information obtained using standard experimental techniques.
Funding: The study is supported by the Russian Science Foundation (grant No. 23-14-00313).

doi 10.18699/bgrs2024-3.1-34

Epigenetic regulation through nucleosome plasticity: insights from MD simulations

Fedulova A.S.[1]*, Motorin N.A.[1], Shariafetdinova A.S.[1], Armeev G.A.[1], Shaytan A.K.[1, 2]

[1] *Department of Biology, Lomonosov Moscow State University, Moscow, Russia*

[2] *Bioinformatics Lab, Faculty of Computer Science, HSE University, Moscow, Russia*

* *a.kniazeva@intbio.org*

Key words: nucleosomes; chromatin; molecular dynamics simulations

Motivation and Aim: Eukaryotic DNA exists within a structure called chromatin. Chromatin and its basic unit, the nucleosome, are highly dynamic structures. The nucleosome core particle is formed by an octamer of histone proteins and 145–147 base pairs of wrapped DNA. Nucleosome dynamics regulate the accessibility of DNA to molecular machines and facilitate nuclear protein interactions. While experimental studies of nucleosome dynamics provide a static snapshots of dynamically stable states, additional methods, such as molecular modeling, are essential to interpret the indirect data. We use molecular dynamics (MD) simulations to describe nucleosome dynamics and mechanisms underlying its epigenetic regulation. Recent rapid improvements in computational capabilities have enabled the modeling of atomistic nucleosome dynamics on a multi-microsecond timescale.

Methods and Algorithms: We used all-atom MD simulations of nucleosomes whose initial structures were derived from the Protein Data Bank. Due to the nucleosome's substantial size and the presence of intrinsically disordered regions, MD simulations are particularly challenging. The study of nucleosomal DNA and protein (histone) globular core dynamics has been carried out on nucleosome X-ray structures with truncated disordered regions (histone tails) and using the TIP3P water model. For the histone tails simulations, different water models had to be tested. MD simulations were performed using the Gromacs, Amber ff14SB with parmbsc1 and CUFIX corrections. The NVT ensemble was used with a reference temperature and pressure of 300 K and 1 bar. Ion concentration of 150 mM NaCl was used.

Results: MD simulations in the multi-microsecond time scale allowed to simulate the partial unwinding of nucleosomal DNA from the histone core (both for nucleosomes with full-length and truncated tails) [1–3]. Additionally, DNA sliding coupled with histone-DNA contact reorganization was observed [2]. Detailed analysis showed that nucleosomal DNA dynamics are accompanied by bending motions of the H2A-H2B histone dimer. Such bending dynamics were characterized by principal component analysis as a general dynamic mode of H2A-H2B dimers (within the nucleosome and intact dimer) [4]. We hypothesize that epigenetic factors, such as histone variant incorporation or peptide binding, influence nucleosome dynamics through these histone core dynamic modes. Further MD simulations demonstrated that nucleosomes containing the variant histone H2A.Z exhibited a greater degree of bending in H2A-H2B dimers compared to canonical nucleosomes [4]. This observation aligns with findings from other studies that reported more pronounced DNA unwinding in variant

nucleosomes [5]. In another example, binding of the viral peptide LANA to the histone surface of the nucleosome stabilizes a more bent H2A-H2B dimer [6].

The influence of histone tail modifications (a major source of epigenetic marks) remains poorly understood. The MD studies should take into account recent findings on the use of various water models. Previously, the study by colleagues presented the comparison of NMR data (R1 and R2 relaxation rates) with the MD-derived analogue parameters and stated that the TIP4P-D water model is the most suitable for such studies [7]. In our latest simulations, the closest values of the R1/R2 relaxation parameters were obtained in the OPC water model.

Conclusion: MD simulations have provided mechanistic insights into the interplay between DNA dynamics and histone core dynamics within nucleosomes. The role of histone bending in DNA sliding and unwinding has been described, as well as how the presence of histone variants or peptide binding can influence nucleosome dynamics. These findings enhance our understanding of the structural and dynamic complexity of nucleosomes, particularly in the context of epigenetic regulation.

Funding and Acknowledgements: The study is supported by the Russian Science Foundation grant #23-74-10012 (https://rscf.ru/en/project/23-74-10012/). The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University.

References

1. Winogradoff D., Aksimentiev A. Molecular Mechanism of Spontaneous Nucleosome Unraveling. *J Mol Biol*. 2019;431(2):323-335. doi 10.1016/j.jmb.2018.11.013
2. Armeev G.A., Kniazeva A.S., Komarova G.A., Kirpichnikov M.P., Shaytan A.K. Histone dynamics mediate DNA unwrapping and sliding in nucleosomes. *Nat Commun*. 2021;12(1):2387. doi 10.1038/s41467-021-22636-9
3. Huertas J., Schöler H.R., Cojocaru V. Histone tails cooperate to control the breathing of genomic nucleosomes. *PLoS Comput Biol*. 2021;17(6):e1009013. doi 10.1371/journal.pcbi.1009013
4. Kniazeva A.S., Armeev G.A., Shaytan A.K. H2A-H2B Histone Dimer Plasticity and Its Functional Implications. *Cells*. 2022;11(18):2837. doi 10.3390/cells11182837
5. Li S., Wei T., Panchenko A.R. Histone variant H2A.Z modulates nucleosome dynamics to promote DNA accessibility. *Nat Commun*. 2023;14(1):769. doi 10.1038/s41467-023-36465-5
6. Oleinikov P.D., Fedulova A.S., Armeev G.A. et al. Interactions of Nucleosomes with Acidic Patch-Binding Peptides: A Combined Structural Bioinformatics, Molecular Modeling, Fluorescence Polarization, and Single-Molecule FRET Study. *IJMS*. 2023;24(20):15194. doi 10.3390/ijms242015194
7. Rabdano S.O., Shannon M.D., Izmailov S.A. et al. Histone H4 Tails in Nucleosomes: a Fuzzy Interaction with DNA. *Angew Chem Int Ed*. 2021;60(12):6480-6487. doi 10.1002/anie.202012046

doi 10.18699/bgrs2024-3.1-36

Neurotrophin receptors: the transmembrane domain as a key player of the receptor activation and a target for antidepressants

Goncharuk S.*, Kot E., Mineev K.
Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia
** ms.goncharuk@gmail.com*

Key words: neurotrophin receptor; Trk; TrkA; TrkB; P75NTR; antidepressants; fluoxetine; NMR

Motivation and Aim: Neurotrophins and their receptors are key players in the development of the vertebrate nervous system and therefore represent a target for the treatment of various neurodegenerative disorders and cancers. The neurotrophins signal through two types of cellular receptors – the Trk family of tyrosine kinases and the p75NTR. To elucidate the mechanism of receptor activation and thus to facilitate rational drug design, the three-dimensional structures are required. To date, the structures of extracellular and intracellular domains of P75 and Trk receptors have been obtained, including ligand-binding state, however, the activation process is still uncovered. The blind spot is the structure of transmembrane domains in different states. Here we present the structures of transmembrane domain of neurotrophin receptors in active and inactive states, and study the interaction of TrkB with antidepressants.
Methods and Algorithms: Cell-free protein synthesis was used to produce the transmembrane domain (TMD) of neurotrophin receptors. The proteins were purified by size-exclusion chromatography and incorporated into detergent micelles or lipid-detergent bicelles for structural studies. To solve the structures of TMD the NMR spectroscopy was employed. Functional assays were used to investigate the biological relevance of these structures.
Results: We solved the structure of TrkA TMD in an inactive state and proposed that the receptor is activated by rotation of the TMD pre-formed dimers upon NGF binding. We show that although extracellular juxtamembrane regions are unstructured, they are likely interact with ligands and thus switch the state of TMD dimer. We also solve the structure of TrkB TMD by NMR in a lipid environment. The biological relevance of this structure was verified by functional assays using mutagenesis. We showed that the conformation found corresponds to an active state of the receptor. We investigated the propensity of TrkB TMD to interact with several antidepressants and found out that drugs stabilize the TrkB TMD dimer.
Conclusion: The spatial structures of Trk TMD presented here make an important branch of the tree of knowledge required to understand the neurotrophin signaling. Our findings on TrkB/drugs interaction confirm that the TrkB TMDs are promising drug targets, and the presented NMR spectroscopy approach can be used to screen new chemical compounds.
Funding: The study was supported by the Russian Science Foundation (project No. 22-14-00130).

doi 10.18699/bgrs2024-3.1-37

New insights into the properly folded membrane protein production for structural studies

Goncharuk M.V.[1]*, Motov V.V.[1], Kot E.F.[1], Kornilov F.D.[1], Savitskaya A.G.[1], Mineev K.S.[1, 2], Arseniev A.S.[1], Goncharuk S.A.[1]

[1] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

[2] *Current address: Institute of Organic Chemistry and Chemical Biology, Goethe University Frankfurt, Frankfurt am Main, Germany*

** m.s.goncharuk@gmail.com*

Key words: membrane protein; transmembrane fragment; protein production; bacterial expression; cell-free synthesis; protein purification; active protein; refolding from organic solvent; protein ligation; membrane mimetic environment

Motivation and Aim: Membrane proteins are involved in many processes in living cells. Their malfunction is associated with diseases, including socially important ones. Despite the significance, the exact way of their function at the molecular level is poorly studied. Partially this takes place because of substantial difficulties in preparation of properly folded samples (i. e. in an active state or in native conformation) for structural and functional investigation. Despite all recent achievements, there is still no general approach to natively folded membrane protein production. To fill this gap in knowledge, here we present three independent strategies: helical membrane protein refolding from organic solvent mixtures, cell-free synthesis of full-length active membrane protein and protein ligation as well as discuss their scope of application.

Methods and Algorithms: Bacterial expression and continuous exchange cell-free synthesis were applied for protein production. Common chromatographic techniques such as immobilised metal affinity chromatography and gel filtration were used for protein purification. A wide set of membrane mimicking media was screened for incorporation of the transmembrane domains of membrane proteins. Protein ligation was performed in an enzyme-depended reactions.

Results: We considered three different approaches to prepare the properly folded membrane proteins. First approach is the refolding of the helical membrane proteins from organic solvent mixtures. A wide range of parameters was screened. Optimal composition of the mixture for lipid/detergent particle formation from the organic-water solutions was established and the effect of the protein on this process was investigated. The rational technique of protein incorporation was developed and approved for several membrane proteins from different classes. Second approach is cell-free synthesis of full-length active membrane protein. An extensive multi-parametric screening of co-translational incorporation of protein into the membrane mimetics (detergent micelles and bicelles, liposomes and lipid-protein nanodiscs of different size and composition) revealed that Facade-based bicelles inhibits translation machinery in much lower extent than other detergent-based mimetics did. The yield of an active membrane protein was comparable to one obtained in the presence of lipid-protein nanodiscs, making Facade-based bicelles a prospective tool to be used for an active membrane protein production. Third approach is the protein ligation. We showed that the hydrophobic transmembrane and hydrophilic intracellular parts of the protein could be produced separately in

different expression systems in properly folded states, and after that successfully ligated into a single molecule by special enzyme. A set of the reaction parameters were tested and discussed.

Conclusion: Three approaches presented here illustrate possible independent ways to obtain the membrane protein samples in native conformation (i.e. in a properly folded state). Despite that each of them on their own cannot be successfully applied to every membrane protein, altogether they significantly expand the spectra of the membrane proteins applicable for structural biology.

Funding: The study is supported by the Russian Science Foundation (grant No. 22-14-00020).

doi 10.18699/bgrs2024-3.1-38

X-ray crystallography in the era of high-quality predictions

Gushchin I.*

Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia
** ivan.gushchin@phystech.edu*

Key words: X-ray crystallography; protein structure; protein engineering; AlphaFold

X-ray crystallography remains the most commonly used experimental method for determining protein structure. Yet, recently, a plethora of machine-learning approaches was introduced that solve almost completely some of the most difficult problems in structural biology. In my talk, I will discuss the recent advances in macromolecular crystallography and computational methods, highlighting the areas where experimental approaches are still indispensable as well as the tasks where the two methods complement each other very well. I will also provide examples where crystallography studies by Russian scientists were instrumental in obtaining important results [1–7].
Funding: The study was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement 075-03-2024-117, project FSMG-2021-0002).

References

1. Lyapina E. et al. Structural Basis for Receptor Selectivity and Inverse Agonism in S1P5 Receptors. *Nat Commun.* 2022,13(1):4736. doi 10.1038/s41467-022-32447-1
2. Nikolaev A., Tropina E.V., Boldyrev K.N., Maksimov E.G., Borshchevskiy V., Mishin A., Yudenko A., Kuzmin A., Kuznetsova E., Semenov O., Remeeva A., Gushchin I. Two Distinct Mechanisms of Flavoprotein Spectral Tuning Revealed by Low-Temperature and Time-Dependent Spectroscopy. *Protein Sci.* 2024;33(1):e4851. doi 10.1002/pro.4851
3. Zhang H. et al. Bright and Stable Monomeric Green Fluorescent Protein Derived from StayGold. *Nat Meth.* 2024;21(4):657–665. doi 10.1038/s41592-024-02203-y
4. Bukhdruker S., Varaksa T., Orekhov P., Grabovec I., Marin E., Kapranov I., Kovalev K., Astashkin R., Kaluzhskiy L., Ivanov A., Mishin A., Rogachev A., Gordeliy V., Gilep A., Strushkevich N., Borshchevskiy V. Structural Insights into the Effects of Glycerol on Ligand Binding to Cytochrome P450. *Acta Cryst D.* 2023;79(1):66-77. doi 10.1107/S2059798322011019
5. Sluchanko N.N.; Slonimskiy Y.B., Egorkin N.A., Varfolomeeva L.A., Kleymenov S.Yu., Minyaev M.E., Faletrov Y.V., Moysenovich A.M., Parshina E.Yu., Friedrich T., Maksimov E.G., Boyko K.M., Popov V.O. Structural Basis for the Carotenoid Binding and Transport Function of a START Domain. *Structure.* 2022;30(12):1647-1659.e4. doi 10.1016/j.str.2022.10.007
6. Sluchanko N.N. et al. Structural Framework for the Understanding Spectroscopic and Functional Signatures of the Cyanobacterial Orange Carotenoid Protein Families. *Int. J. Biol. Macromol.* 2024;254:127874. doi 10.1016/j.ijbiomac.2023.127874
7. Tikhonova T.V., Osipov E.M., Dergousova N.I., Boyko K.M., Elizarov I.M., Gavrilov S.N., Khrenova M.G., Robb F.T., Solovieva A.Y., Bonch-Osmolovskaya E.A., Popov V.O. Extracellular Fe(III) Reductase Structure Reveals a Modular Organization Enabling *S*-Layer Insertion and Electron Transfer to Insoluble Substrates. *Structure.* 2023;31(2):174-184.e3. doi 10.1016/j.str.2022.12.010

doi 10.18699/bgrs2024-3.1-39

PhotoProteinTech database on the properties of luciferases and Ca^{2+}-regulated photoproteins, technologies and areas of their application created by automatic analysis of scientific publications and patents

Ivanisenko V.A.[1, 2, 3, 4]*, Burakova L.P.[4], Eremeeva E.V.[4], Natashin P.V.[4], Larionova M.D.[4], Malikova N.P.[1], Ivanisenko T.V.[1, 2, 3, 4], Demenkov P.S.[1, 2, 3], Venzel A.S.[1, 2, 3, 4], Ivanisenko N.V.[1, 4, 5], Vysotski E.S.[4]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

[4] *Institute of Biophysics, SB RAS, Federal Research Center "Krasnoyarsk Science Center of the Siberian Branch of the Russian Academy of Sciences", Krasnoyarsk, Russia*

[5] *AIRI, Moscow, Russia*

* *salix@bionet.nsc.ru*

A relational database PhotoProteinTech has been developed, containing data on the properties of coelenterazine-dependent luciferases and Ca^{2+}-regulated photoproteins (sequences, physicochemical properties, melting temperatures, emission spectra, and bioluminescent activity), as well as information on technologies and areas of their application.

The database is implemented using MySQL 5.7.26 DBMS and consists of interconnected tables: (i) "main_data" contains data on native luciferases and photoproteins; (ii) "mutants_data" – data on mutant forms of proteins from the "main_data" table; (iii) "homologs_data" – data on homologs of proteins from the "main_data" table; (iv) "func_site_data" – data on functional sites in the protein structure; (v) "ligand_data" – data on molecular ligands; (v) "taxonomy" – taxonomy; (vi) "dictionaries" – dictionaries of names and synonyms of proteins and ligands; (vii) "ReferenceCross" – references to literature; (viii) "References" – description of literature; (ix) "Patents" – description of patents.

To populate the database, an analysis of scientific publications and patents was carried out using natural language text mining methods implemented in our previously developed cognitive software and information system ANDSystem [1]. The current version of the database contains information about the light-sensitive Ca^{2+}-regulated photoprotein berovin from *Beroe abyssicola* and its 78 mutants, Ca^{2+}-regulated photoproteins aequorin and its 96 mutants and obelin and its 87 mutants from *Aequorea victoria* and *Obelia longissima*, respectively, as well as the coelenterazine-dependent Renilla luciferase and its 130 mutants.

Analysis of international patents made it possible to extract information about a wide range of technologies in which photoproteins and luciferases were used, including "BRET-based antibody sensors", "luminescent antibody sensing (LUMABS) technology", "bioluminescence resonance energy transfer assays (BRET)", "NanoBRET technology", "High-throughput screening", "in vivo imaging" technologies, etc., as well as their biomedical applications, including areas such as drug resistance, cardiovascular diseases, tumors, and others.

In conclusion, PhotoProteinTech provides data on the physicochemical and structural properties of luciferases and photoproteins, as well as related information on technologies and areas of their practical application, which determines the novelty of the created database. The presented data can be used to design new forms of luciferases and photoproteins.

Funding: The work was supported by the grant of the Russian Science Foundation No. 22-14-00125 “Rational design of coelenterazine-dependent bioluminescent proteins with novel properties using mutagenesis and molecular modeling”.

References

1. Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics.* 2019;20(Suppl. 1):H.34

doi 10.18699/bgrs2024-3.1-40

Interaction energy model of SARS-CoV-2 M^{pro} with peptides in cleavage sites

Kadtsyn E.D.[1, 2, 3]*, Kolybalov D.S.[1, 2], Zubavichus Y.V.[1, 2]

[1] *SRF "SKIF", Koltsovo, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Voevodsky Institute of Chemical Kinetics and Combustion, SB RAS, Novosibirsk, Russia*

* *e.kadtsyn@g.nsu.ru*

Key words: SARS-CoV-2 M^{pro}; drug design; peptidomimetics; molecular dynamics; MMPBSA

Motivation and Aim: SARS-CoV-2 main protease (M^{pro}) is a common target for design of anti-COVID drugs due to its high conservativity, difference from human proteases and critical role in virus lifecycle. A tempting strategy for viral protease inhibition is a use of peptidomimetics. To design a peptidomimetic with high inhibition constant an information about peptide-enzyme complex structure and interaction energy is highly desirable. It is difficult to obtain crystallographic structure of such complexes due to their metastability. Despite that recently a crystal structures of 11 M^{pro}-peptide complexes were registered, each of peptides consist of 6 residues after the cleavage position with one mutation heightening its stability in complex. Being inspired by that work we carried out a molecular dynamic (MD) simulation of these 11 hexapeptides without mutations as well as of 11 hexapeptides consisted of 3 residues before cleavage position and 3 residues after it. Obtained models were used to estimation of peptide-protease interaction energy and decompose them to energies of each residue. The data were used then to check correlation between energies for different residues and built energy model of interaction.

Methods and Algorithms: Crystallographic structure of M^{pro} was taken from PBD database (7NG6), starting configurations of peptides were generated by AmberTools package from known sequence. 100ns MD simulation of peptides in aqueous solution were carried out to relax the structures, then using cluster analysis the most represented structures were extracted and docking was used to obtain structures of complexes. 500ns MD simulation of each complex were carried out. RMSD, R_g and SASA values were calculated to check an enzyme stability during simulation and influence of peptide to protease. MMPBSA approach was used to estimate interaction energies and per-residue contributions. H-bonds were analyzed to explain energy results.

Results: RMSD, R_g and SASA values for all systems show that complex is stable during the simulation in each case, no notable differences between pure protease and complexes were detected proving no influence of peptide to protease structure. Interaction energies and their contributions were determined. For interaction energies sets of the same amino acids in the same positions *t*-tests and *F*-tests were carried out, no statistically notable differences were shown, independent residue model hence can be applied to calculate peptide energy. Per-residue values for all amino acids presented in studied peptides were collected. From the data the most interacting peptide sequence was hypothesized.

Conclusion: From the model data of amino acid energy contributions one can construct the structure of highly-interacting peptide as well as highly-interacting peptidomimetic. Data of H-bonds between each peptides residue and protease was used to explain results.

The data obtained should be used in future to rationally design peptidomimetic drugs suitable for anti-COVID therapy.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation within the governmental order for SRF SKIF Boreskov Institute of Catalysis (project FWUR-2024-0040)

References

1. Lee J. et al. X-ray crystallographic characterization of the SARS-CoV-2 main protease polyprotein cleavage sites essential for viral processing and maturation. *Nat Commun*. 2022;13(1):5196. doi 10.1038/s41467-022-32854-4

doi 10.18699/bgrs2024-3.1-41

Sequence-based computational protein analysis of the MDR pump AcrAB-TolC from *Escherichia coli*

Karakozova M.[1, 2]*, Raldugina V.[1], Nazarov P.A.[1]
[1] *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*
[2] *All-Russian Scientific Research Institute of Medicinal and Aromatic Plants, Moscow, Russia*
* *mvk752002@gmail.com*

Key words: alignment; MDR pump; AcrAB-TolC; sequence

Motivation and Aim: Multidrug-resistance (MDR) pumps play a vital role in safeguarding bacteria against antibiotics. MDR pumps provide a foundation for non-specific protection of bacteria and are classified into six families of MDR pumps. One of the most extensively researched MDR pumps is the *Escherichia coli* AcrAB-TolC pump [1]. *Methods and Algorithms*: Reference sequences of proteins TolC, AcrA and AcrB for the *E. coli* MG1655 were found and downloaded from the NCBI database in the formats genbank and fasta using the biopython package and Etrez.Esearch functions and Entrez.Efetch. To download sequences, analyze and align them the biopython package was used. Sequences were aligned between yourself using BLAST to create a custom database [2].
Results: Studying pump sequences is crucial for comprehending the development of new clinical isolates or superbugs, as efflux-related resistance plays a significant role among the total number of bacteria that become resistant due to antibiotic treatment. The genes responsible for encoding MDR pumps are seen as adaptable and are categorized as "luxury" genes, meaning they help bacteria adapt to changing environmental conditions. Examination of laboratory strains revealed that the sequences of many commonly used laboratory strains (K-12 and B) remain unchanged and show no mutations in the amino acid sequences. The conservation of amino acid sequences in the pumps mirrors the conservation seen in housekeeping genes, underscoring the significant role of these genes in bacteria, even though they are classified as nonessential genes [3].
Were analyzed 8383 TolC protein sequences, 6239 AcrA protein sequences, 3561 AcrB protein sequences. We showed that almost every amino acid is variable.
Conclusion: Having analyzed all known sequences of the AcrAB-TolC pump proteins, we found that the use of a bioinformatics approach to sequence analysis leads to incorrect results due to the quality of the database. Almost every sequence required preliminary validation, which changes the analysis pipeline.
Funding: This work was supported by grant from the Russian Science Foundation 22-15-00099 (Creation of a heterologous MDR pump system for studying bacterial resistance of pathogenic gram-negative bacteria).

References

1. Nazarov P.A., Kuznetsova A.M., Karakozova M.V. Multidrug Resistance Pumps as a Keystone of Bacterial Resistance. *Moscow Univ Biol Sci Bull*. 2022;77(4):193-200. doi 10.3103/S009639252204006X
2. Raldugina V. Sequences of TolC-containing complexes of MDR transporters of various gram-negative bacteria analysis. Diploma work, 2023
3. Karakozova M.V., Nazarov P.A. Conserved sequences of genes coding for the multidrug resistance pump AcrAB-TolC of Escherichia coli suggest their involvement into permanent cell "cleaning". *Bull RSMU*. 2018;2:32-36

doi 10.18699/bgrs2024-3.1-42

Binding free energy of small molecule/GLP-1R complexes calculations with funnel metadynamics

Krivosheev A.S.[1]*, Golovin A.V.[1, 2]

[1] *Sirius University of Science and Technology, Sochi, Russia*

[2] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

** artemkr0804@yandex.ru*

Key words: funnel metadynamics; molecular dynamics; GLP-1R; small molecules; binding free energy; diabetes mellitus; obesity

Motivation and Aim: The glucagon-like peptide-1 (GLP-1) hormone is released following food intake [1] and plays a pivotal role in regulating glucose levels within the body. Upon secretion, GLP-1 interacts with its receptor, GLP-1 receptor (GLP-1R), triggering a cascade of events that includes the stimulation of insulin secretion and the inhibition of glucagon release [2]. This hormonal pathway not only aids in maintaining glucose homeostasis but also influences other metabolic processes. Activation of GLP-1R by GLP-1 leads to various physiological responses, including the delay of gastric emptying [3], heightened feelings of satiety [4], suppression of food intake, and consequent weight reduction [5]. Consequently, harnessing the effects of GLP-1 and its receptor has emerged as a promising therapeutic strategy for managing conditions such as Type 2 Diabetes Mellitus (T2DM) and obesity.

Currently, multiple peptidic GLP-1R agonists, including semaglutide and liraglutide, have received approval for the treatment of T2DM [6, 7, 8]. However, the administration of them via subcutaneous injection poses challenges in patient adherence and convenience, highlighting the need for more accessible delivery methods. Oral administration emerges as a preferable alternative, offering improved compliance and convenience [9]. Moreover, emerging evidence suggests that GLP-1R agonists not only address metabolic dysregulation but also confer cardiovascular benefits by reducing blood pressure [10] and enhancing microvascular function. However, the development of small molecule drugs targeting GLP-1R activation poses significant challenges [11].

The exploration of ligand-receptor interactions often relies on assessing binding free energy. However, common methods for this analysis are resource-intensive and time-consuming, limiting their practicality for precise results. We propose utilizing funnel metadynamics [12], a computational technique that enhances sampling efficiency by employing a funnel-shaped restraint potential. This approach guides ligands through relevant conformational space, reducing unnecessary exploration of irrelevant regions.

Our aim is to investigate the binding free energy of small molecules with the GLP-1R using funnel metadynamics. We aim to validate this method's accuracy using molecules with known binding free energies as ground truth. Furthermore, we seek to explore novel molecules as potential candidates for developing therapeutics targeting T2DM and obesity.

Methods and Algorithms: Atomic partial charges were calculated using PsiRESP [13], which employs the Psi4 quantum chemistry engine. Molecular topologies were generated using ACPYPE, incorporating charges obtained from PsiRESP and the GAFF2 force field.

Molecular dynamics simulations were employed to analyze the temporal movements of atoms and molecules. Additionally, funnel metadynamics was utilized to restrict available space during simulation. This technique also enabled enhanced sampling of binding free energy space by employing collective variables, specifically the distance along the cone axis and the distance from the cone axis.

Results: Our study has yielded results demonstrating close agreement between the calculated values of free binding energy and experimental data, validating the accuracy of used approach. Furthermore, we have determined free binding energies for novel small molecules interacting with the GLP-1R, offering valuable insights into their potential as therapeutic candidates. The results will be used further in the ongoing development of drugs that activate GLP-1R.

Conclusion: Our study has shown that the funnel metadynamics method is effective. By using a funnel-shaped restraint potential, we were able to focus more precisely on exploring the binding site of the receptor and ligand, improving our understanding of their interaction. This suggests that funnel metadynamics could be valuable for future research aimed at identifying potential agonist molecules for the receptor we studied. This approach could lead to better ways of studying how drugs interact with their targets, which is crucial for discovering new treatments in the field of molecular pharmacology. Further investigation using funnel metadynamics may help us uncover complex molecular interactions and develop new drugs that are more effective and specific in their actions.

References

1. Orskov C., Holst J.J., Knuhtsen S., Baldissera F.G., Poulsen S.S., Nielsen O.V. Glucagon-like peptides GLP-1 and GLP-2, predicted products of the glucagon gene, are secreted separately from pig small intestine but not pancreas. *Endocrinology*. 1986;119:1467-1475. doi 10.1210/endo-119-4-1467
2. Kreymann B., Williams G., Ghatei M.A., Bloom S.R. Glucagon-like peptide-1 7-36: a physiological incretin in man. *Lancet*. 1987;2(8571):1300-1304. doi 10.1016/s0140-6736(87)91194-9
3. Nauck M.A., Niedereichholz U., Ettler R., Holst J.J., Orskov C., Ritzel R., Schmiegel W.H. Glucagon-like peptide 1 inhibition of gastric emptying outweighs its insulinotropic effects in healthy humans. *Am J Physiol*. 1997;273(5):E981-E988. doi 10.1152/ajpendo.1997.273.5.E981
4. Turton M.D., O'Shea D., Gunn I. et al. A role for glucagon-like peptide-1 in the central regulation of feeding. *Nature*. 1996;379(6560):69-72. doi 10.1038/379069a0
5. Drucker D.J. Mechanisms of Action and Therapeutic Application of Glucagon-like Peptide-1. *Cell Metab*. 2018;27(4):740-756. doi 10.1016/j.cmet.2018.03.001
6. Nauck M. Incretin therapies: highlighting common features and differences in the modes of action of glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors. *Diabetes Obes Metab*. 2016;18(3):203-216. doi 10.1111/dom.12591
7. Novo Nordisk A/S. WEGOVY (semaglutide) prescribing information. https://www.novo-pi.com/wegovy.pdf (accessed 2024-05-15)
8. Novo Nordisk A/S. SAXENDA (liraglutide) prescribing information. https://www.novo-pi.com/saxenda.pdf (accessed 2024-05-15)
9. Dibonaventura M.D., Wagner J.S., Girman C.J. et al. Multinational Internet-based survey of patient preference for newer oral or injectable Type 2 diabetes medication. *Patient Prefer Adherence*. 2010;4:397-406. doi 10.2147/PPA.S14477
10. Wen S., Nguyen T., Gong M., Yuan X., Wang C., Jin J., Zhou L. An Overview of Similarities and Differences in Metabolic Actions and Effects of Central Nervous System Between Glucagon-Like Peptide-1 Receptor Agonists (GLP-1RAs) and Sodium Glucose Co-Transporter-2 Inhibitors (SGLT-2is). *Diabetes Metab Syndr Obes*. 2021;14:2955-2972. doi 10.2147/DMSO.S312527
11. de Graaf C., Song G., Cao C., Zhao Q., Wang M.W., Wu B., Stevens R.C. Extending the Structural View of Class B GPCRs. *Trends Biochem Sci*. 2017;42(12):946-960. doi 10.1016/j.tibs.2017.10.003
12. Limongelli V., Bonomi M., Parrinello M. Funnel metadynamics as accurate binding free-energy method. *Proc Natl Acad Sci USA*. 2013;110(16):6358-6363. doi 10.1073/pnas.1303186110
13. Wang L., O'Mara M.L. PsiRESP: calculating RESP charges with Psi4. *J Open Source Software*. 2022;7(73):4100. doi 10.21105/joss.04100

doi 10.18699/bgrs2024-3.1-43

Comprehensive analysis of photosensitizer-biomolecule interactions using hybrid EPR and computational methods

Krumkacheva O.[1]*, Kolokolov M.[1, 2], Sannikova N.[1], Dementyev S.[1, 2], Podarov R.[1, 2], Fedin M.[1]

[1] *International Tomography Center, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *olesya@tomo.nsc.ru*

Key words: Photosensitizers; Protein-Ligand Interactions; EPR; Molecular Modeling

Motivation and Aim: Photodynamic therapy (PDT) is a promising non-invasive therapeutic approach valued for its selectivity. PDT relies on photosensitizers (PSs) that accumulate at tumor sites and interact with oxygen and light to produce reactive oxygen species. These reactive species induce controlled cell death and enhance immune responses. PSs are non-toxic until activated by light, distinguishing PDT from other cancer treatments. However, the effectiveness of PDT is influenced by the biodistribution and pharmacodynamics of the PSs.

The intracellular localization of PSs directly impacts their mechanism of action and overall efficacy in PDT. Localization within target tissues and cell organelles depends on the chemical structure of PSs, including charge, lipophilicity, size, and asymmetry. Most clinically used PSs are macrocyclic tetrapyrrole structures (such as porphyrins, chlorines, and phthalocyanines) administered intravenously. These compounds are chosen for their high absorption in the visible light spectrum and their efficiency in generating singlet oxygen. However, their tendency to aggregate and their poor tumor-targeting capabilities limit their potential in PDT.

One strategy to enhance PDT effectiveness is using transport proteins to improve PS solubility and distribution. Human serum albumin (HSA), the most abundant protein in blood plasma, serves as a natural drug carrier due to its biocompatibility, water solubility, and ability to bind various compounds. HSA's hydrophobic pockets can stabilize PSs, preventing aggregation and extending their half-life. The affinity of PSs for HSA affects their plasma concentration, distribution, and therapeutic efficacy. However, the exact binding sites and interactions between PSs and HSA remain poorly understood, complicating efforts to optimize PSs for clinical use.

Traditional methods like fluorescence-based techniques provide limited and often ambiguous information about PS binding sites on HSA. These methods struggle to distinguish between multiple binding sites. Pulsed dipolar Electron Paramagnetic Resonance (EPR) spectroscopy provides nanometer-scale structural information and captures conformationally distinct states of the system under study without limitations on biomolecule size. Unlike other methods, pulsed dipolar EPR spectroscopy provides the whole distance distribution between labels.

Recently, LaserIMD (laser-induced magnetic dipole) spectroscopy has been proposed to measure the distances between the photoexcited state of PS and a stabilized nitroxide label. Ongoing methodological work on light-induced PDS using photoexcited triplet states opens new opportunities for structural studies of biological complexes with photoactive molecules, including those used in PDT [1]. LaserIMD EPR spectroscopy

offers precise measurements of distances between PSs and nitroxide spin labels on protein, capturing the full distance distribution between labels. This technique overcomes the limitations of other methods by discriminating between different binding sites. However, even with precise distance measurements, identifying exact binding sites can be challenging due to complex structural dynamics.

The aim of this study [2] is to develop and apply this integrative approach to investigate the interactions between various PSs (anionic, neutral, and cationic) and HSA. By combining high-resolution EPR spectroscopy with sophisticated computational models, we aim to provide a detailed, experimentally validated map of PS binding sites on HSA. This will enhance our understanding of PS-HSA interactions and support the development of more effective PDT agents.

Methods and Algorithms: In this study, we employed a novel integrative approach combining experimental laser-induced EPR spectroscopy with advanced computational techniques, including molecular docking and molecular dynamics simulations, to investigate the structures of human serum albumin with photosensitizers.

The following steps outline our methodology:

1. Laser-Induced Dipolar EPR Spectroscopy (LaserIMD). The experimental investigation began with LaserIMD EPR spectroscopy conducted at 30 K using a Q-band Bruker Elexsys E580 spectrometer equipped with a cryogenic system from Oxford Instruments. Samples were photoexcited using a Nd:YAG laser at 532 nm. This method provided high-resolution distance measurements between nitroxide spin labels and photoexcited porphyrins, revealing intricate details of the binding sites and conformational dynamics of the complexes.
2. Molecular Docking. Initial blind docking was performed to identify potential binding sites of photosensitizers on HSA. Utilizing AutoDock-GPU, we conducted extensive docking simulations with 1000 LGA runs on a population of 2048 individuals, leveraging GPU acceleration to manage computation times efficiently. The use of GPU acceleration allowed for the generation of up to 1000,000 poses, significantly enhancing the thoroughness of the search.
3. Cluster Analysis and Filtering. After the initial docking, the generated poses were subjected to k-means clustering to group similar binding poses. Clusters were filtered by comparing the potential binding sites with experimental distance distributions obtained from LaserIMD. This rough comparison helped eliminate clusters that were clearly inconsistent with experimental data, ensuring that only plausible binding sites were considered for further analysis.
4. Focused Docking. The filtered clusters from the blind docking phase were subjected to focused docking. This step involved refining the predictions by concentrating on specific regions of interest identified during the initial docking phase. Smaller grid boxes tailored for each type of photosensitizer were employed, enhancing the accuracy of the binding site predictions.
5. Molecular Dynamics (MD) Simulations. The docked complexes underwent molecular dynamics simulations using GROMACS with the Gromos54a7 force field. Ligand topologies were generated by the Automated Topology Builder (ATB). The system was subjected to energy minimization, followed by heating and equilibration, before a 100 ns production run. MD simulations accounted for the mobility and dynamic behavior of the complexes, ensuring that the modeled distances reflected realistic conformational changes.

6. Spin Label Conformation Modeling. To accurately model the distance distributions between spin labels and photosensitizers, we used the ChiLife software to predict the possible conformations of spin labels attached to the Cys34 residue.
7. Comparison and Validation. The experimentally obtained distance distributions from LaserIMD were compared with those predicted by molecular dynamics simulations. This detailed comparison ensured that the computational models accurately reflected the experimental data. The integration of experimental EPR measurements with computational docking and dynamics provided a robust framework for identifying and validating binding sites.

Results: The integrative approach combining LaserIMD EPR spectroscopy and molecular modeling provides clear structural information, including precise distances between spin labels at Cys34 of HSA and porphyrins. This approach allows for accurate filtering and comparison of molecular modeling results with experimental data, enabling the construction of a comprehensive, experimentally validated picture of photosensitizer binding sites and their distribution.

Our findings reveal that binding does not always occur in the standard sites of albumin and often involves multiple sites, which is crucial for interpreting fluorescence data commonly used for characterizing binding. Additionally, comparing docking results with experimental data highlights the limitations of blind docking without experimental validation, necessitating a reevaluation of previous results.

We identified the influence of substituent charge and the presence of metal ions on the localization of binding sites. The consistency between calculated and experimental distance distributions validates our approach.

This methodology is not limited to proteins and can be applied to other systems of interest. For instance, similar techniques were used to investigate the binding of quadruplex DNA with the porphyrin TmPyP4 [3].

Conclusion: In summary, the results demonstrate the potential of our developed approach for investigating biological complexes with photoactive ligands. Combining direct observation of photoactive centers via EPR with the predictive power of computational modeling offers a powerful tool for addressing challenges in photodynamic therapy and other biological applications.

Funding: The work was supported by Russian Science Foundation (20-73-10239).

References

1. Sannikova N.E., Timofeev I.O., Chubarov A.S. et al. Application of EPR to Porphyrin-Protein Agents for Photodynamic Therapy. *J Photochem Photobiol B*. 2020;211:112008. doi 10.1016/j.jphotobiol.2020.112008
2. Sannikova N.E. et al. Multicenter EPR-Based Approach for the Localization of Photosensitizers in Biomolecules. In: Physics and Chemistry of Elementary Chemical Processes. Proceedings of the 10th International Voevodsky Conference. Novosibirsk, 2022;162
3. Sannikova N.E., Kolokolov M.I., Khlynova T.A. et al. Revealing Light-Induced Structural Shifts in G-Quadruplex-Porphyrin Complexes: A Pulsed Dipolar EPR Study. *Phys Chem Chem Phys*. 2023;25(33):22455-22466. doi 10.1039/D3CP01775C

doi 10.18699/bgrs2024-3.1-44

The influence of zinc ions on the spatial organization and activation of TIR-domains of the Toll-like receptors

Lushpa V.[1, 2]*, Goncharuk M.[1], Lin C.[3], Talyzina I.[1#], Luginina A.[2], Vakhrameev D.[2], Shevtsov M.[2], Goncharuk S.[1, 2], Arseniev A.[1], Borchevskiy V.[2, 4, 5], Wang X.[3, 6], Mineev K.[1†]

[1] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

[3] *Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, China*

[4] *Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich GmbH, Jülich, Germany*

[5] *JuStruct: Jülich Center for Structural Biology, Forschungszentrum Jülich GmbH, Jülich, Germany*

[6] *Department of Applied Chemistry and Engineering, University of Science and Technology of China, Hefei, China*

 # Current address: Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA

† Current address: Institute of Organic Chemistry and Chemical Biology, Goethe University Frankfurt, Frankfurt am Main, Germany

** lushpa1696@gmail.com*

Key words: toll-like receptors; solution-state NMR; molecular modeling; spatial structure; Zn binding; DoE; competitive binding

Motivation and Aim: Proteins of the toll-like receptor (TLR) family are the key components of the human innate immune system [1, 2]. There are 10 different TLRs (TLR1-10) present in the human body. Activation of these receptors initiates the launch of the innate immune response and the inflammatory process. It has been established that TLRs are involved to varying degrees in the development of infectious, autoimmune and neurodegenerative diseases [3].

TLRs have a characteristic type 1 transmembrane protein structure, including large extracellular and intracellular domains, as well as a single transmembrane alpha-helical region. These receptors function in the form of homo- and heterodimers. Despite numerous studies of TLRs, many questions remain unresolved regarding the structural organization of the intracellular domains of TLRs and the mechanism of their activation during signal transduction to activate the immune response.

Methods and Algorithms: To select the conditions for the production of all TLR1-10 proteins, the experimental design method was used, described in detail in the articles [4, 5]. The development of cleaning protocols was based on previously obtained results. For TLR1 TIR (Toll/Il-1 Receptor)-domain, protein production and purification is described in detail in the article [6]. The selected parameters were used for subsequent two-step purification using IMAC and SEC followed by sample concentration. Validation of the secondary structure of proteins was carried out using CD spectrometry and NMR.

Triple resonance spectra were used to determine the main and side chains of TLR1 TIR by NMR. The data obtained were used to calculate the spatial structure automatically with manual correction in areas with increased mobility of the protein chain. Data from

NMR relaxation experiments were used to study the dynamic properties of the protein. More details about the research methodology can be found in the article [7].
Experiments examining the metal binding activity of TLR TIR domains were performed using NMR and analyzed in Mathematica [8]. The research methodology using the example of TLR1 TIR is published in the article [7]. Functional tests of the studied receptors are also described in the article [7].
Results: The conditions for the production and purification of TLR TIR-domains were selected. Based on the data obtained, candidates for structural studies were selected. Structure and dynamics data in solution were obtained for TLR1 TIR. Studies of the metal-binding activity of TLR TIRs have revealed the formation of protein-zinc complexes with nanomolar binding constants. In experiments with point mutants of TIR domains of TLRs, amino acid residues necessary for binding to zinc ions were identified in the sequences of the proteins under study. The results of functional analysis of TLR are demonstrated, according to which the role of zinc ion binding in TLR activation is shown, and the functional significance of previously identified key amino acid residues is presented. The results obtained for TLR1 TIR were published in [7].
Conclusion: Differences have been established between the spatial structures of TLR1 TIR obtained in solution and crystal. The zinc binding activity of intracellular TLR domains was revealed and the functional significance of zinc on the activation of the TLR1/2 receptor was demonstrated. The key role of a number of amino acid residues of TLR TIRs in receptor activation has been revealed. The data obtained significantly add to the knowledge of the functioning of TLR receptors, and in particular, suggest that the zinc-binding ability of the TLR1 TIR-domain is critical for receptor activation.
Funding: The study was supported by the Russian Science Foundation grant No. 22-14-00020.

References

1. Fitzgerald K.A., Kagan J.C. Toll-like Receptors and the Control of Immunity. *Cell*. 2020;180(6):1044-1066
2. Medzhitov R. Toll-like receptors and innate immunity. *Nat Rev Immunol*. 2001;1(2):135-145
3. Hammerich L., Marron T.U., Upadhyay R. et al. Systemic clinical tumor regressions and potentiation of PD1 blockade with in situ vaccination. *Nat Med*. 2019;25(5):814-824
4. Draper N.R., Davis T.P., Pozueta L., Grove D.M. Isolation of Degrees of Freedom for Box–Behnken Designs. *Technometrics*. 1994;36(3):283-291
5. Gutiérrez-González M., Farías C., Tello S. et al. Optimization of culture conditions for the expression of three different insoluble proteins in Escherichia coli. *Sci Rep*. 2019;9(1):16850
6. Goncharuk M.V., Lushpa V.A., Goncharuk S.A. et al. Sampling the cultivation parameter space for the bacterial production of TLR1 intracellular domain reveals the multiple optima. *Protein Expr Purif*. 2021;181:105832
7. Lushpa V.A., Goncharuk M.V., Lin C. et al. Modulation of Toll-like receptor 1 intracellular domain structure and ac-tivity by Zn2+ ions. *Commun Biol*. 2021;4(1):1003
8. Wolfram Research, Inc. Mathematica / Wolfram Research, Inc. Wolf-ram Research, Inc., 2024

doi 10.18699/bgrs2024-3.1-45

Protein-interacting C2H2-type zinc fingers: structural studies by NMR

Mariasina S.[1]*, Dukhalin S.[1], Balagurov K.[2], Efimov S.[3], Bocharov E.[4], Bonchuk A.[2], Polshakov V.[1]

[1] *Lomonosov Moscow State University, Moscow, Russia*

[2] *Institute of Gene Biology, Moscow, Russia*

[3] *Kazan Federal University, Kazan, Russia*

[4] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow, Russia*

* *sm1024sm@yandex.ru*

Key words: zinc finger C2H2 type; protein NMR; transcription factors; CLAMP; MSL2; CG18262

Motivation and Aim: Transcription factors are proteins that regulate gene expression. One common type of domain found in transcription factors is the C2H2 zinc finger. This type of domain has two β-strands and an α-helix, which is stabilized by a zinc ion coordinated by two cysteine and two histidine residues. These domains are typically involved in specific DNA binding through the residues in the α-helix. However, there is increasing evidence that these domains may also play a role in protein-protein interactions.

In our study, we investigated the structural features and protein-protein interactions of two N-terminal C2H2 zinc finger domains from different *Drosophila* proteins. The first studied object was the Chromatin-Linked Adaptor for MSL Proteins (CLAMP). Previous research has shown that CLAMP directly interacts with the male-specific organizer of the complex (MSL2) [1]. MSL2 plays an important role in the specificity of binding to the X chromosome during the process of dosage compensation in *Drosophila*.

The second object is CG18262. Based on our preliminary investigation, it appears to be a member of a potentially new family of C2H2 zinc fingers. Unlike typical domains of this type, members of this new family have an additional β-strand at the N-terminus, which we hypothesize could be an effector domain responsible for dimerization.

Structural data are essential for understanding the physiological functions and interactions between proteins. This study aims to determine the three-dimensional solution structures of Zn finger protein domains and investigate their protein-protein interaction networks. The findings of this research will contribute to a better understanding of the processes involved in transcription regulation.

Methods: Heteronuclear NMR spectroscopy was used to obtain structural information. Unlabeled, ^{15}N-labeled, and ^{15}N/^{13}C-labeled protein samples were expressed in *E. coli*. 2D and 3D NMR spectra were recorded using NMR spectrometers with 600 and 700 MHz frequency, equipped with either a room temperature triple resonance probe or a cryoprobe.

Resonance assignments were made using a standard 3D heteronuclear NMR technique. Restrained molecular dynamics protocol was applied to obtain 3D structure of the proteins, using the distance restraints derived from nuclear Overhauser effects, torsion angle restraints obtained from the analysis of NMR chemical shifts, and hydrogen bond

restraints measured in H/D exchange NMR experiments. Protein-protein interactions were studied using the NMR titrations and isotope filtered experiments.
Results: We obtained the NMR solution structure of the N-terminal C2H2 Zn finger domain of the CLAMP protein (residues 87–153, 7.4 kDa). This domain has a classical C2H2 zinc finger fold with a rather unusual distribution of residues typically used for DNA recognition (PDB entry 7NF9, Fig. 1A).

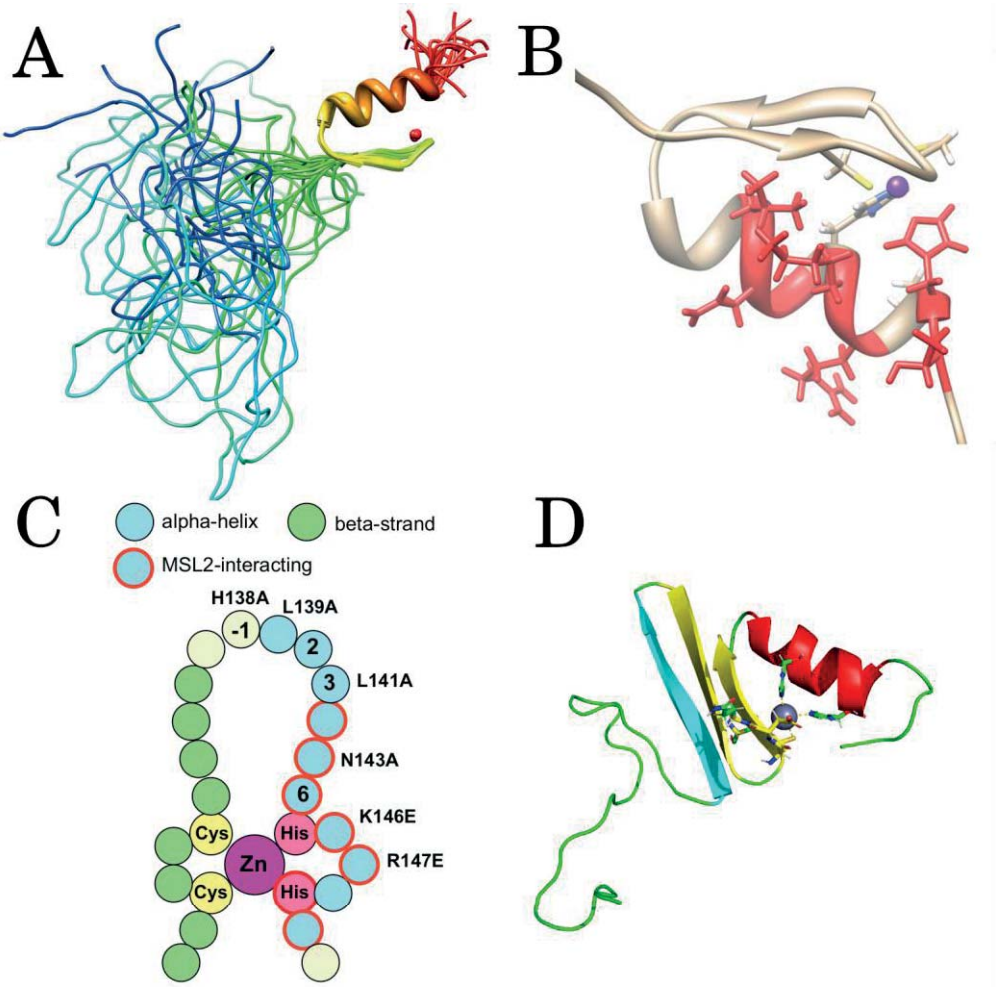


Fig. 1. Structural studies of N-terminal zinc finger domain of CLAMP and CG18262. **A**. Family of NMR solution structures of CLAMP$^{87\text{-}153}$ (PDB id 7NF9). **B**. 3D structure of CLAMP$^{87\text{-}153}$ with residues interacting with MSL2 shown in red. **C**. Schematic representation of the secondary structure of the zinc finger domain of CLAMP, with α-helical residues shown in blue and β-strand residues in green. are represented by sticks. The MSL2-binding residues, which were the subject of mutagenesis, are surrounded by red circles. **D**. Cartoon representation of the NMR solution structure of CG18262$^{1\text{-}62}$. The β-strands from the canonical C2H2 zinc finger are shown in yellow, while residues 25-32, which form an additional β-strand responsible for dimerization, are shown in cyan

To identify the critical residues for MSL2 binding pattern within CLAMP, we carried out NMR titration experiments using ^{15}N-labeled CLAMP$^{40-153}$ or CLAMP$^{1-153}$ in the presence of an excess of unlabeled MSL2$^{618-655}$. These experiments showed that only residues within the α-helix of C2H2 Zn finger are involved in interactions with MSL2 (Fig. 1B and C). The roles of these residues were confirmed by mutagenesis experiments. The dissociation constant (K_d) for the CLAMP–MSL2 complex was determined to be relatively weak (0.2 ± 0.1 mM at 25 °C), but the interactions were highly specific [2]. We also obtained the structure of the N-terminal domain of CG18262 (residues 1-62, Fig. 1D). As expected, it appears to be a typical C2H2 zinc finger, but it is stabilized by an additional β-strand at the N-terminus.

Consistent with our model predictions, the N-terminal domain of CG18262$^{1-62}$ has dimerization activity, which was confirmed by NMR experiments using mixtures of ^{15}N,^{13}C-labbeled and unlabeled protein. The dimerizing interface involves residues 25-32 attributed to the additional β-strand in the C2H2 zinc finger.

Conclusion: In this study, structural data on two protein-interacting N-terminal domains of the C2H2 zinc fingers from *Drosophila* have been obtained. CLAMP$^{87-153}$ is a classical C2H2 zinc finger with an unusual ability to interact with its protein partner MSL2 via set of residues on its α-helix. The second protein, CG18262, has an unusual structure for a canonical C2H2 zinc finger, with an additional β-strand at the N-terminus that represents an effector domain responsible for dimerization.

Funding: The study is supported by the Russian Science Foundation (№24-14-00081).

References

1. Tikhonova E., Fedotova A., Bonchuk A., Mogila V., Larschan E.N., Georgiev P., Maksimenko O. The simultaneous interaction of MSL2 with CLAMP and DNA provides redundancy in the initiation of dosage compensation in drosophila males. *Development*. 2019;146(19):dev179663. doi 10.1242/dev.179663
2. Tikhonova E., Mariasina S., Efimov S., Polshakov V., Maksimenko O., Georgiev P., Bonchuk A. Structural basis for interaction between CLAMP and MSL2 proteins involved in the specific recruitment of the dosage compensation complex in Drosophila. *Nucleic Acids Res.* 2022;50(11):6521-6531. doi 10.1093/nar/gkac455

doi 10.18699/bgrs2024-3.1-46

Revealing G-quadruplex DNA structures in HPV16 using quantitative PCR stop and ligands

Martin C.[1, 2], Calderón R.[1], Nahuat N.[1], Conde L.[1], González M.R.[1], Kantún N.[1]*

[1] *Centro de Investigaciones Regionales 'Dr. Hideyo Noguchi', Yucatan, Mexico*

[2] *Universidad Anáhuac Mayab, Yucatan, Mexico*

* *nuvia.kantun@correo.uady.mx*

Key words: G-quadruplexes; DNA; HPV 16; qPCR stop; ligands

Motivation and Aim: G-quadruplex helical structures (G4) represent non-canonical conformations that can emerge in guanine-rich regions of nucleic acids under specific intracellular conditions, such as the presence of monovalent cations, pH fluctuations, temperature changes, and cellular crowding. Additionally, they can be modulated by small molecules, also known as G4 ligands [1]. The non-random occurrence of G4 structures within viral genomes accentuates their relevance in various biological contexts. The identification of G4 structures in diverse viral genomes of medical importance suggests their potential roles as pivotal actors in infectious mechanisms, positioning them as promising therapeutic targets [2]. Recently, over 100 putative G-quadruplex-forming sequences (PQS) have been predicted in the genomes of three high-risk Human Papillomaviruses using bioinformatic tools [3]. HPV are primary causative agents of the most prevalent sexually transmitted infection worldwide. Genotype 16 is highlighted due to its high risk for oncogenicity; however, it lacks direct antiviral treatments [4, 5]. Accordingly, exploring the structural features of the HPV 16 genome and understanding their role could reveal strategies for designing antiviral treatments against this infection. Therefore, this research aims to confirm the *in vitro* presence of G-quadruplex DNA structures to lay the foundation for further studies of their functions and possible application as molecular targets. These findings could provide significative insights in virology and contribute to the search for new therapeutic approaches for viral infections, including HPV.

Methods and Algorithms: First, we utilized data from the G4-forming sequences or PQS database, as reported by [3], and classified it into clusters based on their genomic position, DNA strand, and G4 score prediction. Subsequently, specific primer sets were designed to amplify G4 clusters located within the E6, E7, E1, E2, L1, and L2 genes, as well as within the long control region or promoter (LCR) of the HPV16 genome. Additionally, primer pairs were designed to flank non-G4 regions, which were predicted by bioinformatic programs in the HPV16 genome, serving as a control region for qPCR stop application. To validate the technique, plasmid DNA (pDNA) containing a known DNA-G4 sequence reported for HPV52 (ID PDB: 5O4D) in the LCR region was utilized as a reference. For this purpose, the HPV52 LCR region (1,163 bp) was cloned in DH5-α cells, as it contains a confirmed G4 by NMR (23 bp), providing a reference for standardization purposes. Primer pairs were designed to flank the reference G4 sequence in HPV52 and a predicted non-G4 sequence in the same genotype, serving as an amplification control. Optimal thermocycling conditions were determined for each primer pair for both HPV16 and 52 by qPCR. Once established, the concentration of KCl was standardized and applied in reactions with different concentrations of the ligands

PhenDC3, PDS, and BRACO-19 to assess their stabilizing effect on G4-DNA, after which three ligand concentrations were selected. These same conditions were applied for G4-DNA detection in HPV 16. The qPCR stop reactions were performed using the GoTaq SYBR PCR master mix (PROMEGA) and ran in a CFX96 Deep Well PCR detection system (BIORAD). A clone containing the complete genome of HPV 16 was used as template to detect the presence of G4-DNA in almost all evaluated regions of HPV 16 genome, except for the ORF L1, which was cloned to supplement the existing biological material. The obtained C_t values were collected and processed with the modified $2^{-\Delta\Delta Ct}$ formula to obtain the relative inhibition degree exerted by the stabilizing effect of ligands on the G4s. Subsequently, a one-way analysis of variance (ANOVA) was conducted to determine significant differences between the ligand concentrations and their effect on each evaluated G4 region. Finally, Dunnett's multiple comparison test was performed to determine differences between the assessed concentrations.

Results: PQS data were classified into 23 clusters distributed in the HPV16 genome, with 12 on the plus strand and 11 on the minus strand. The most abundant genes with PQS were L2 (62), followed by E2 (47) and E1 (36). Al least five PQS clusters were selected for their *in vitro* validation, with PQS scores ≥0.7. Two bacterial clones were obtained: one with the LCR region of HPV52, containing a confirmed G4, and another holding a fragment of the L1 gene of HPV16, serving as DNA templates for the qPCR stop assay. Annealing temperatures ranging from 62 to 66 °C were established for the HPV16 primer pairs corresponding to G4 regions in the early and late genes, the LCR region, non-G4 regions, and those designed for standardization with HPV52. Regarding the technique, the chosen concentration of KCl was 35 mM, based on the evidence obtained during the standardization process, demonstrating its ability to induce the formation of the confirmed G4 of HPV 52 without affecting the detection system.

Additionally, other studies identifying G4s use of concentrations between 25 to 50 mM, supporting its use within this range. Notably, the selected potassium value is lower than that reported in normal cellular environments (100–150 mM) and in carcinogenesis conditions (60 mM), where the formation of a greater quantity of G4s is promoted [6]. Ligand concentrations were determined, resulting in 25, 50 and 100 nM for PhenDC3, while 10, 50 and 100 nM were chosen for PDS and BRACO-19. The assay successfully confirmed the presence of G4-DNA in the E6, E7, E2, E1, L1, and L2 genes of HPV16, with the mentioned ligands exhibiting differential reactivity.

Conclusions: A qPCR Stop assay was standardized at 35 mM KCl and various concentrations of the ligands PhenDC3, PDS, and BRACO-19 to confirm G4-DNA structures in HPV16, resulting in different responses when exposed to PhenDC3, PDS and BRACO-19 ligands. The G4 region located into E6 ORF stood out for its sensibility towards the evaluated ligand concentrations. The confirmed G4 structures could play a relevant role in viral infection cycles and pose an interesting opportunity for future research and antiviral treatment development.

Funding: This study was supported by CONAHCYT through the Frontier Science Grant (CF-2023-I-600): “Evidenciando G-cuádruples en el virus del papiloma humano tipo 16: Una nueva visión estructural de su ADN como blanco terapéutico” under the technical management of Dr. Nuvia Kantún Moreno, professor and researcher at the Regional Research Center ‘Dr. Hideyo Noguchi’.

References

1. Jana J., Mohr S., Vianney Y.M., Weisz K. Structural motifs and intramolecular interactions in non-canonical G-Quadruplexes. *RSC Chem Biol*. 2021;2(2):338-353. doi 10.1039/d0cb00211a
2. Abiri A., Lavigne M., Rezaei M., Nikzad S., Zare P., Mergny J.L., Rahimi H.R. Unlocking G-Quadruplexes as Antiviral Targets. *Pharmacol Rev*. 2021;73(3):897-923. doi 10.1124/pharmrev.120.000230
3. Ontiveros N.J. Caracterización in silico de estructuras G-cuádruples en los genomas de los Virus del Papiloma Humano 16, 18 y 58. Tesis de Licenciatura. Universidad Autónoma de Yucatán, 2023
4. Khairkhah N., Bolhassani A., Najafipour R. Current and future direction in treatment of HPV-related cervical disease. *J Mol Med*. 2022;100(6):829-845. doi 10.1007/s00109-022-02199-y
5. Zheng K., Egawa N., Shiraz A., Katakuse M., Okamura M., Griffin H.M., Doorbar J. The Reservoir of Persistent Human Papillomavirus Infection; Strategies for Elimination Using Anti-Viral Therapies. *Viruses*. 2022;14(2):214. doi 10.3390/v14020214
6. Tateishi-Karimata H., Kawauchi K., Sugimoto N. Destabilization of DNA G-Quadruplexes by Chemical Environment Changes during Tumor Progression Facilitates Transcription. *J Am Chem Soc*. 2018;140(2):642-651. doi 10.1021/jacs.7b09449

doi 10.18699/bgrs2024-3.1-47

Structural studies of the point mutant S25C of the gram-negative targeting endolysin LysSi3 with broad bactericidal activity

Matyuta I.O.[1]*, Sluchanko N.N.,[1] Vasina D.V.[2], Boyko K.M.[1]

[1] *Institute of Biochemistry named after A.N. Bach, Research Center of Biotechnology RAS, Moscow, Russia*

[2] *N.F. Gamaleya National Research Centre for Epidemiology and Microbiology, Ministry of Health of the Russian Federation, Moscow, Russia*

* *i.matyuta@fbras.ru*

Key words: X-ray crystallography; endolysin; point mutant; bacteriophage

Motivation and Aim: Antimicrobial resistance (AMR) is alarmingly increasing in medicine and is one of the major concerns of healthcare today. Due to this fact there is a growing interest in bacteriolytic enzymes that act to lyse bacterial cell wall especially against pathogens with AMR. These types of enzyme include bacteriophage-encoded endolysins which degrade the peptidoglycan (PG) cell wall polymer [1]. Endolysins are synthesized in the cytoplasm of infected bacteria at the final stage of their lytic development for specific degradation the PG polymers of the host bacteria. It leads to abrupt osmotic cell lysis and subsequent release of progeny phages. This phenomenon led to a growing number of studies aimed at using this class of enzymes as antibacterial agents. However, the mechanism of PG binding and degradation by endolysins is still obscure. The LysSi3 is a peptidoglycan hydrolyzing, lysozyme-type enzyme with predicted muramidase activity and broad bactericidal activity. LysSi3 was shown to be a monomer in solution. However, a dimer molecule is observed in the structure of the wild type LysSi3. Examination of the dimer interface revealed that the side chains of S25 of the adjacent subunits are in close proximity. To identify the ability of enzyme to form a dimer it was suggested to obtain and solve the structure of LysSi3 point mutant S25C ($LysSi3^{S25C}$) using X-ray crystallography.

Methods and Algorithms: Initial crystallization screening was performed on a robotic system (Rigaku Americas Corporation, The Woodlands, TX USA) using 96-well VDX plates (Hampton Research, Aliso Viejo, CA USA) and commercial crystallization screens from Hampton Research (Aliso Viejo, CA USA) and Molecular Dimensions Inc (Holland, OH USA) by the "hanging drop" vapor diffusion method. A 15 mg/mL of the $LysSi3^{S25C}$ in 30 mM Tris-HCl buffer pH 8.0 containing 150 mM NaCl was mixed with the crystallization solution in the ratios 1:1, 1:2 (0.1 μL drop volume), and 2:1 (0.2 μL drop volume). The volume of the precipitant solution in the reservoir was 50 μL. The initial crystallization hit was observed under the following conditions: 0.1 M Imidazole pH 7.0 and 20 % v/v Jeffamine ED-2001 pH 7.0 at 1:1 ratio at 288 K.

Crystal of the $LysSi3^{S25C}$ was briefly soaked in a mother liquor containing 20 % glycerol immediately before diffraction data collection and flash-frozen in liquid nitrogen. Dataset was collected at 100K at BL18U beamline (SSRF, China). The dataset was indexed and integrated using the XDS package [2] and scaled using Aimless program of the CCP4 suite [3]. Space groups were suggested by Pointless [4] as $C222_1$.

The structure of LysSi3^{S25C} was solved by the molecular replacement method using the MOLREP program [5] with the atomic coordinates of wild type LysSi3 as a starting model. One copy of the protein was found in an asymmetric unit. The refinement was carried out using the REFMAC5 program of the CCP4 suite [3]. The visual inspection of electron density maps and the manual rebuilding of the model were carried out using the COOT interactive graphics program [6]. The isotropic B-factor and the hydrogen atoms in fixed positions were included during the refinement.
The visual inspection of the modeled structure was carried out using the COOT program and the PyMOL Molecular Graphics System, Version 4.6 (Schrödinger, USA). The contacts were analyzed using the PDBePISA [7].
Results: Structure of LysSi3^{S25C} was solved at 1.8 Å resolution. There is one protein subunit in the asymmetric unit in LysSi3^{S25C} but a dimer similar to LysSi3 was found in the crystal. A double position of the loop with C25 with half occupancy was observed in the electron density. Therefore, there are two possible mutual arrangements of cysteines. In the first case, a disulfide bond is formed. In the second case it is not formed, but then a distance of 2.1 Å is formed between the backbone oxygen atoms of C25 of the adjacent subunits. As a result, the structure shows the formation of a disulfide bond between the C25 of adjacent subunits. Using SEC-MALS and gel electrophoresis it was shown that under oxidized conditions LysSi3^{S25C} is a dimer and in the presence of a reducing agents LysSi3^{S25C} is a monomer.
Conclusion: LysSi3 in solution can form the dimer found in the structures of wild type and point mutant S25C.
Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (No. 075-15-2021-1354).

References

1. Gerstmans H., Rodríguez-Rubio L., Lavigne R., Briers Y. From endolysins to Artilysin®s: novel enzyme-based approaches to kill drug-resistant bacteria. *Biochem Soc Trans*. 2016;44(1):123-128. doi 10.1042/BST20150192
2. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 2010;66:125-132. doi 10.1107/S0907444909047337
3. Winn M.D., Ballard C.C., Cowtan K.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr*. 2011;67:235-242. doi 10.1107/S0907444910045749
4. Evans P. Scaling and assessment of data quality. *Acta Crystallogr Sect D Biol Crystallogr*. 2006;62:72-82. doi 10.1107/S0907444905036693
5. Vagin A.A., Isupov M.N. Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr Sect D Biol. Crystallogr*. 2001;57:1451-1456. doi 10.1107/S0907444901012409
6. Emsley P., Lohkamp B., Scott W.G., Cowtan K. Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr*. 2010;66:486-501. doi 10.1107/S0907444910007493
7. Krissinel E., Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372:774-797. doi 10.1016/j.jmb.2007.05.022

doi 10.18699/bgrs2024-3.1-48

Thermal stability of the rhodopsin of Antarctic UV resistant bacteria

Okhrimenko I.S.*, Lyubaykina N.A., Zagryadskaya Yu. A.
Research Center for Molecular Mechanisms of Ageing and Age-related Diseases of Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia
** okhrimenko.is@mipt.ru*

Key words: rhodopsins; ion pumps; ion channels; X-ray crystallography; photocycle

Motivation and Aim: The most abundant rhodopsin family is the family of proteorhodopsins (PRs)[1]. As many microbial rhodopsins PRs are often used by host organisms to survive in extreme environment. PRs genes were firstly discovered in the marine uncultured proteobacteria and characterised *in vitro*[2]. Later the PRs genes were found in plankton *Archaea*, where they probably appeared due to the lateral gene transfer[3]. The reported PRs have proton pump[2,4,5] or sensory activity[6], they are spread all over the Earth and have the major impact to the solar energy capturing in the world ocean[7]. The PRs genes were found even in Antarctic Sea ice samples[8] (moderate UV areas) and in *Modern Stromatolites* living a High-Altitude Andean Lake in Argentinean Puna in the extreme UV irradiation areas[9]. According to the data of modelling rhodopsin named HbR1 may bind carotenoids. Its gene together with the HbR2 gene were found trough bioinformatics research[10] in the genome of the Antarctic UV resistant bacteria *Hymenobacter sp.* PAMC 26554. This bacterium was isolated from an Antarctic lichen *Usnea sp.*[11] at Barton Peninsula, King George Island, Antarctica (62°13'S, 58°47'W).
It is the open question, if mentioned rhodopsins itselves have special properties originated from Antarctic environment, and what properties are important for the global ecology. We hope that our functional and structural studies will contribute to the answer on these question.
Conclusion: The gene optimized for the best yields of HbR1 was expressed in *E. coli* in presence of all-trans retinal, the purified protein had absorption maximum at 516 nm. HbR1 has not show any light-induced ion transporting activity and had remarkably long photocycle. The length of photocycle is approximately the same at the grate range of H^+ concentrations (pH4.2-9.5) and in the low or hight ion strength of the solution (0 or 1M NaCl). This could show the absence of the connection of proton acceptor group (which should be close to the retinal Shiff base, as analogue of Asp 85 in bacteriorhodopsin of *H. salinarum*) with the environment. HbR1 is stable protein: by means of TSA it's melting temperature was determined as ~80℃, and it increases in case of lowering pH and growing of ion strength. So stability of the protein increases at low pH and high salt conditions.
Funding: The research was performed with the support of Russian Scientific Foundation, project # 23-14-00160.

References

1. Bamann C., Bamberg E., Wachtveitl J., Glaubitz C. Proteorhodopsin. *Biochim Biophys Acta*. 2014;1837(5):614-625
2. Béjà O., Aravind L., Koonin E.V. et al. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*. 2000;289(5486):1902-1906

3. Frigaard N.U., Martinez A., Mincer T.J., DeLong E.F. Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature*. 2006;439(7078):847-850
4. Petrovskaya L.E., Balashov S.P., Lukashev E.P. et al. ESR – A retinal protein with unusual properties from Exiguobacterium sibiricum. *Biochemistry (Mosc)*. 2015;80(6):688-700
5. Yoshizawa S., Kawanabe A., Ito H., Kandori H., Kogure K. Diversity and functional analysis of proteorhodopsin in marine Flavobacteria. *Environ Microbiol*. 201214(5):1240-1248
6. Hartz A.J., Sherr B.F., Sherr E.B. Photoresponse in the heterotrophic marine dinoflagellate Oxyrrhis marina. *J Eukaryot Microbiol*. 2011;58:171-177
7. Gómez-Consarnau L. et al. Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Sci Adv*. 2019;5(8):eaaw8855
8. Koh E.Y. et al. Proteorhodopsin-bearing bacteria in Antarctic sea ice. *Appl Environ Microbiol*. 2010;76(17):5918-5925
9. Albarracín V.H. et al. Functional green-tuned proteorhodopsin from modern stromatolites. *PLoS One*. 2016;11(5):e0154962
10. Ushakov A., Grudinin S., Okhrimenko I. et al. Knowledge-based prediction model for characterization of microbial rhodopsins for optogenetics. *FEBS J*. 2016;283:127-128
11. Bubach D. et al. Elemental composition of Usnea sp lichen from Potter Peninsula, 25 de Mayo (King George) Island, Antarctica. *Environ Pollut*. 2016;210:238-245

doi 10.18699/bgrs2024-3.1-49

Investigation of the rotor effect of modified 8-oxo-adenosine as part of DNA duplexes

Pushkarevskaya A.A.[1, 2]*, Aralov A.V.[3], Lomzov A.A.[1, 2]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

* *lomzov@niboch.nsc.ru*

Key words: rotor effect; fluorescent oligonucleotide probes; hybridization; DNA duplexes; SNP

Motivation and Aim: The detection of single nucleotide polymorphisms (SNPs) in nucleic acids is an urgent task of modern biomedicine. One of the approaches allowing their detection is the creation of highly specific fluorescent oligonucleotide probes. The main goal of our research is to test the physicochemical and structural properties of DNA duplexes containing a previously unexplored fluorophore: 8-oxo-N6-((3-methyl)-2-benzothiazolone)-2′-deoxyadenosine (X). The main purpose of my study is to gain an understanding of whether this fluorophore can be used to search for SNPs.

Methods and Algorithms: The fluorescence properties of X in the free state, as part of oligonucleotides and their duplexes, were determined by quantum yield determination using a standard. The hybridization properties of X were evaluated by thermal denaturation and computer modeling methods.

Results: The physicochemical properties of (8-oxo-6-(3-methyl-2-benzothiazolone)-adenosine and oligodeoxyribonucleotides containing this fluorescent nucleoside were investigated as follows. The fluorescence quantum yield value of the nucleotide was 1.2 ± 0.2 % fluorescence. The value of quantum yield of oligonucleotides containing X increased from 9 to 22 times. When forming duplexes with DNA, changes in the value of fluorescence the quantum yield are multidirectional, depending on the nucleotide environment of the fluorophore. Our hypothesis was not confirmed since we did not see significant changes in X fluorescence during duplex formation. The results of my study indicate that as part of duplexes, the fluorophore cannot become brighter because its rotation around the C6-N bond is too mobile regardless of duplex composition. There is a need to redesign or restructure our research in a different way.

The thermal stability of complexes with single-nucleotide mismatches increases in the series of bases G > T > C; destabilization of complexes is observed only in the case of the location of A opposite the modification. In complexes in which the modified nucleotide is surrounded by pyrimidine bases and adenine is opposite the modification, the greatest destabilization of DNA duplexes is observed.

Since we tested a large number of X-containing sequences, we hypothesize that changing the design of a conventional fluorophore will help improve its fluorescent properties in DNA duplexes.

Conclusion: Today, a large number of fluorophores have been developed; however, debate continues about the best strategies for the design and improvement of the properties of such molecules.

Most studies in the field of the development and research of fluorophores have only focused on creating simple and cheap molecules to synthesize. But the results of previous

studies have proved inconclusive because they strongly disturb the DNA structure or have insufficient fluorescence intensity. To sum up, it is possible to confirm the existing theory that the quantum yield of oligonucleotides and their duplexes increases significantly in comparison with the free nucleoside.

Funding: The research was performed within the framework of the state assignment ICBFM SB RAS No. 123021600208-7.

doi 10.18699/bgrs2024-3.1-50

Understanding nucleosome dynamics and interactions through integrative approaches

Shaytan A.K.[1, 2]*

[1] *Lomonosov Moscow State University, Moscow, Russia*

[2] *Bioinformatics Lab, Faculty of Computer Science, HSE University, Moscow, Russia*

* *shaytan_ak@mail.bio.msu.ru*

Key words: nucleosome; structural biology; integrative modeling; molecular dynamics; histones

Nucleosome are elementary building blocks of chromatin that wrap DNA around an octamer of histone proteins. They not only compact the genomic DNA in eukaryotes but also intricately participate in all genomic processes including gene expression regulation, epigenetic memory, and development [1]. Nucleosomes bridge the "digital" nature of the genetic information stored in the DNA sequence with the "analog" world of physical interactions between the nucleic acids and proteins used to interpret this genetic information. Nucleosome dynamics, its conformational flexibility, and its dependence on nucleosome composition and environmental factors is an important pathway used to regulate genome functions. Understanding nucleosome structure and dynamics, as well as its functional implication, requires a complex integrative approach which combines molecular modeling, experimental, and bioinformatics analysis.

Methods and Algorithms: To understand nucleosome dynamics and interactions we have employed a combination of experimental, molecular modeling, and bioinformatics approaches. These include all-atom molecular dynamics simulations, DNA footprinting techniques, small angle X-ray scattering, solution NMR, fluorescence-based assays, en masse analysis of the available X-ray and cryo-EM structures, analysis of genomics data to identify histone protein variants and mutations [2–8].

Results: Our recent reanalysis of the human genome and the available gene expression data allowed us to update the current nomenclature of the histone genes for the human genome, suggesting that there are collectively 90 protein-coding core histone genes (histone H3, H4, H2A, H2B) that may be incorporated into nucleosomes conferring different properties through sequence variations [2]. Some of these genes are often mutated in cancer, with mutations affecting nucleosome stability or the functioning of important histone post-translational modification sites [3]. We have shown that all-atom molecular dynamics simulations at a multi-microsecond timescale may reveal new functional dynamics modes within nucleosomes that are responsible for DNA unwrapping and sliding [4]. The analysis of nucleosomes incorporating different histone variants and mutations with MD simulations confirmed that the intrinsic dynamics of nucleosomes are affected by such variations [5]. We have also shown that nucleosome dynamics is affected by their interaction with nucleosome-binding peptides and proteins [6]. A comprehensive analysis of all available PDB structures of nucleosomes conducted by us suggests that certain modes of nucleosome dynamics can be characterized by such high-throughput comparison [7, 8]. Still, such processes as DNA unwrapping and histone tail dynamics remain poorly characterized. Our efforts in analyzing nucleosomes with small-angle X-ray scattering suggest that these methods can be used to understand the effects of DNA sequence on nucleosome compactness. The characterization of the

flexible histone tails is possible through the application of solution NMR techniques. The measurement of the NMR relaxation parameters (R1 and R2) may be used to assess the histone tail dynamics and validate different molecular dynamics force fields.

Conclusion: We have shown that the characterization of nucleosome structure and dynamics, especially, its dependence on its composition and interactions requires a complex integrative approach uniting different experimental methods, molecular modeling methods, and bioinformatics methods. However, further progress is needed, especially focused on our ability to characterize and predict the dynamics of nucleosomes with varying composition, including variations in DNA sequence, histone sequence, histone PTMs, and DNA modifications.

Funding: The study is supported by the Russian Science Foundation (grant No. 19-74-30003).

References

1. Armeev G.A., Gribkova A.K., Pospelova I., Komarova G.A., Shaytan A.K. Linking Chromatin Composition and Structural Dynamics at the Nucleosome Level. *Curr Opin Struct Biol*. 2019;56:46-55. doi 10.1016/j.sbi.2018.11.006
2. Seal R.L., Denny P., Bruford E.A. et al. A Standardized Nomenclature for Mammalian Histone Genes. *Epigenet Chromatin*. 2022;15:34. doi 10.1186/s13072-022-00467-2
3. Espiritu D., Gribkova A.K., Gupta S., Shaytan A.K., Panchenko A.R. Molecular Mechanisms of Oncogenesis through the Lens of Nucleosomes and Histones. *J Phys Chem B*. 2021;125(16):3963-3976. doi 10.1021/acs.jpcb.1c00694
4. Armeev G.A., Kniazeva A.S., Komarova G.A., Kirpichnikov M.P., Shaytan A.K. Histone Dynamics Mediate DNA Unwrapping and Sliding in Nucleosomes. *Nat Commun*. 2021;12(1):2387. doi 10.1038/s41467-021-22636-9
5. Kniazeva A.S., Armeev G.A., Shaytan A.K. H2A-H2B Histone Dimer Plasticity and Its Functional Implications. *Cells*. 2022;11:2837. doi 10.3390/cells11182837
6. Oleinikov P.D., Fedulova A.S., Armeev G.A. et al. Interactions of Nucleosomes with Acidic Patch-Binding Peptides: A Combined Structural Bioinformatics, Molecular Modeling, Fluorescence Polarization, and Single-Molecule FRET Study. *IJMS*. 2023;24:15194. doi 10.3390/ijms242015194
7. Armeev G.A., Gribkova A.K., Shaytan A.K. NucleosomeDB – a database of 3D nucleosome structures and their complexes with comparative analysis toolkit. *bioRxiv*. 2003. doi 10.1101/2023.04.17.537230
8. Armeev G.A., Gribkova A.K., Shaytan A.K. Nucleosomes and Their Complexes in the cryoEM Era: Trends and Limitations. *Front Mol Biosci*. 2022;9:1070489. doi 10.3389/fmolb.2022.1070489

doi 10.18699/bgrs2024-3.1-51

Details of the structure and function of the bacterial thiocyanate dehydrogenase with the unique copper active site

Varfolomeeva L.A.[1]*, Shipkov N.S.[1], Dergousova N.I.[1], Boyko K.M.[1], Tikhonova T.V.[1], Popov V.O.[1, 2]

[1] *Federal Research Centre "Fundamentals of Biotechnology", RAS, Moscow, Russia*

[2] *Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia*

** larisaavarfolomeeva@gmail.com*

Key words: thiocyanate dehydrogenase; trinuclear copper center; active site rearrangements; complex with inhibitors; point mutagenesis; investigation of catalytic mechanism; X-ray crystallography

Motivation and Aim: The oxidoreductases adapted the copper as a cofactor to catalyze oxidation or reduction of the wide spectrum of low and high molecular weight substrates providing the special environment for the metal ions [1]. Revealing the relationships between the structure and function of the enzyme copper centers can be useful for biotechnological applications. Thiocyanate dehydrogenase (TcDH) is the enzyme participating in a particular metabolic pathway of thiocyanate oxidation and containing the trinuclear copper center with unique architecture [2]. The aim of the present work was to characterize the structure and function of the TcDH trinuclear copper center in detail.

Methods and Algorithms: X-ray crystallography was applied as a main method in the present work to investigate the TcDH structural details, including trinuclear copper center. pmTcDH from a bacterium *Pelomicrobium methylotrophicum* was used as object of investigation in the present work.

Results: Crystals of pmTcDH turned out to scatter X-ray at near atomic resolution that allowed to shed light on fine details of the enzyme active site. pmTcDH is a homodimer with each subunit folded as a seven bladed β-propeller. The main particular feature of the pmTcDH structure is different conformations of the subunits within one dimer: one subunit has a closed active site, while the other is accessible for solvent. The hypothesis of the cooperativity between subunit of the enzyme dimer was refused by the structural data for the apo pmTcDH, where such relationship was not observed.

The pmTcDH active site is located in a central cavity of a β-propeller and includes three copper ions Cu1, Cu2, Cu3, coordinated by six histidine residues as well as D279 and K68 residues. Switching from the open to the closed conformation accompanies with the peculiar active site changes. Observed rearrangements of the active site are supposed to be key steps of the first stages of a catalytic reaction, such as the substrate binding and the deprotonating conservative water molecule.

To verify a model of the substrate orientation in the active site the structures of the pmTcDH complexes with two competitive inhibitors, thiourea and substrate analogue selenocyanate were solved. Both inhibitors are observed to bound only in the closed pmTcDH active site. The obtained structural data confirms the model of the thiocyanate binding proposed in [2].

Conclusion: The rearrangements of the TcDH active site during the catalytic reaction are characterized in detail based on near-atomic resolution structural data. The model of the

substate binding is confirmed based on the structure of the enzyme complexes with inhibitors. Obtained data shed light on the mechanism of a catalytic reaction of TcDH.
Funding: The work is supported by the Russian Science Foundation grant #23-74-30004.

References

1. Solomon E.I., Heppner D.E., Johnston E.M. et al. Copper active sites in biology. *Chem Rev.* 2014;114(7):3659-3853. doi 10.1021/cr400327t
2. Tikhonova T.V., Sorokin D.Y., Hagen W.R. et al. Trinuclear copper biocatalytic center forms an active site of thiocyanate dehydrogenase. *Proc Natl Acad Sci USA.* 2020;117(10):5280-5290. doi 10.1073/pnas.1922133117

doi 10.18699/bgrs2024-3.1-52

Structure of SARS-CoV-2 receptor-binding domain in complex with the virus-neutralizing nanobody sheds light on mechanism of possible viral avoidance of antibody therapy

Varfolomeeva L.A.[1]*, Sluchanko N.N.[1]*, Shcheblyakov D.B.[2]*, Favorskaya I.A.[2], Dolzhikova I.V.[2], Korobkova A.I.[2], Alekseeva I.A.[2], Esmagambetov I.B.[2], Derkaev A.A.[2], Prokofiev V.V.[2], Zorkov I.D.[2], Logunov D.Y.[2], Gintsburg A.L.[2], Popov V.O.[1, 3], Boyko K.M.[1, 2]*

[1] *Federal Research Centre "Fundamentals of Biotechnology", RAS, Moscow, Russia*

[2] *National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya, Ministry of Health of the Russian Federation, Moscow, Russia*

[3] *Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia*

** l.varfolomeeva@fbras.ru; nikolai.sluchanko@mail.ru; sdmitriyv@mail.ru; kmb@inbi.ras.ru*

Key words: X-ray crystallography; protein-protein interaction; neutralizing nanobody; S-protein receptor-binding domain; SARS-CoV-2

Motivation and Aim: The therapeutic use of virus neutralizing antibodies is a prospective approach against some infectious diseases [1], including COVID-19, which has caused millions of deaths worldwide and affected global economics. The single-chain antibody (nanobody) P2C5 targeting the SARS-CoV-2 receptor-binding domain (RBD) has demonstrated virus-neutralizing activity against earlier described variants and been included as a main component in "GamCoviMab". However, the later emerged Delta or Omicron XBB.1 variants turned to be neutralized by the P2C5 nanobody with a decreased effectiveness. The aim of the present work was to reveal an epitope of the P2C5 nanobody and a mechanism of diminishing virus-neutralizing activity in relation to the certain virus variants.

Methods and Algorithms: X-ray crystallography was applied as a main method to elucidate the structure of SARS-CoV-2 RBD fragment complexed with the P2C5 nanobody.

Results: The N-truncated RBD fragment 319–541 complexed with the P2C5 nanobody ($RBD_{319-541}$:P2C5) produced many diverse crystallization condition, which were sensitive to X-ray impact and demonstrated poor quality of a diffraction pattern [2]. Extensive search of better diffracted crystals using a synchrotron source succeeded in the structure of $RBD_{319-541}$:P2C5 at 3.10 Å resolution that was used for further analysis. The interface between RBD and P2C5 with approximately area 940 Å estimated by PISA [3] is stabilized by many polar contacts and hydrophobic interactions. Among hydrophobic residues of RBD involved in an interface formation the residue L452 deserves special attention since namely that residue is mutated to arginine in the Delta variant. The mutation L452R of RBD leads to the electrostatic repulsion with K96 residue of P2C5 that explain diminishing of P2C5 binding to RBD of the Delta variant. As in case of the Omicron XBB.1 variant, the mutation F490S significantly contribute to destabilization of the RBD:P2C5 interface. Being a part of a hydrophobic cluster, the residue F490 is directly stacked with the F47 of P2C5, and the substation of an aromatic moiety to serine with small polar chain disrupts the hydrophobic interaction.

The obtained structural data also demonstrate the putative mechanism of the virus-neutralizing activity of P2C5. The binding of P2C5 with its epitope sterically prevents a complex formation with ACE2 receptor taking place during a virus invasion.
Of note, AlphaFold2/AlphaFold3 both completely failed to proper model RBD:P2C5 complex, indicating the importance of experimental methods for solving such problems.
Conclusion: The obtained structures of RBD:P2C5 complex enable to formulate a putative mechanism of the virus-neutralizing activity of the P2C5 nanobody to earlier SARS-CoV-2 variants and to explain its decreased binding to the Delta and Omicron XBB.1 variants. In addition, the knowledge of the epitope location facilitates a reengineering of P2C5 for the other virus variants and improves its specificity.
Funding: The work is supported by the Russian Science Foundation grant #23-74-30004.

References

1. Hammitt L.L., Dagan R., Yuan Y. et al. Nirsevimab for Prevention of RSV in Healthy Late-Preterm and Term Infants. *New Engl J Med*. 2022;386:837-846. doi 10.1056/NEJMoa2110275
2. Boyko K.M., Varfolomeeva L.A., Egorkin N.A. et al. Preparation and Crystallographic Analysis of a Complex of SARS-CoV-2 S-Protein Receptor-Binding Domain with a Virus-Neutralizing Nanoantibody. *Crystallogr Rep*. 2023;68(6):864-871. doi 10.1134/S1063774523601168
3. Krissinel E., Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372:774-797. doi 10.1016/j.jmb.2007.05.022

doi 10.18699/bgrs2024-3.1-53

Establishing the relationship between SOD1 protein mutations and ALS patient lifespan using molecular modeling and graph neural networks

Venzel A.S.[1, 2]*, Ivanisenko V.A.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *venzel@bionet.nsc.ru*

Key words: SOD1; amyotrophic lateral sclerosis; molecular modeling; graph neural networks

Motivation and Aim: Mutations in the SOD1 gene, encoding the enzyme superoxide dismutase 1, are a leading cause of the hereditary form of amyotrophic lateral sclerosis (ALS), an incurable neurodegenerative disorder. Symptom management is currently the sole therapeutic approach for patients with ALS. Each SOD1 mutation correlates with a specific lifespan for ALS patients. It is proposed that misfolding of the SOD1 protein due to mutations may underlie the disease at the molecular level. These mutations can alter the hydrogen bond network within the protein structure, resulting in mutant conformations distinct from the native protein fold of wild-type. Investigating the changes in hydrogen bonds in mutant variants and their relationship to protein conformation and properties is therefore crucial. Graph neural networks (GNNs), which are becoming popular in structural bioinformatics, can be leveraged to analyze hydrogen bond networks of mutants, characterize SOD1 mutant structures, and identify quantitative relationships between hydrogen bonds and protein properties, such as ALS patient lifespan.

This study aims to establish a connection between the structures of mutant SOD1 protein variants and the lifespan of ALS patients using molecular modeling and graph neural networks.

Methods and Algorithms: The structures of wild-type and mutant SOD1 proteins were obtained using ESMFold. Molecular modeling of SOD1 mutants was carried using the Rosetta macromolecular modeling suite. Graph neural networks were developed using the PyTorch Geometric (PyG) library for Python.

Results: A novel method was developed to correlate protein structure with properties in case of SOD1 protein. The protein structure is represented by a hydrogen bond graph, where nodes represent amino acids and edges denote the stability of hydrogen bonds. This hydrogen bond graph was used to train a graph neural network. The trained network is used to gain vector representations of hydrogen bonds graph of mutants, which were subsequently used to build a regression model to predict the lifespan of ALS patients.

Funding: The study is supported by the budget project, FWNR-2022-0020.

3 Симпозиум «Структурная биология и фармакология: компьютерные и экспериментальные подходы»

Symposium “Structural biology and pharmacology: computational and experimental approaches”

doi 10.18699/bgrs2024-3.2-01

Извлечение знаний о биологической активности лекарственных препаратов: исследование на примере соединений с противовирусными свойствами

Бизюкова Н.*, Соболев Б., Карасев Д., Ионов Н., Сухачёв В., Такташов Р., Рудик А., Иванов С., Тарасова О.

Институт биомедицинской химии им. В.Н. Ореховича, Москва, Россия

** nad.smol@gmail.com*

Ключевые слова: биологическая активность; интеллектуальный анализ текстов; противовирусные препараты; распознавание наименований объектов; извлечение ассоциаций

Мотивация и цель: Важным источником знаний, необходимых во многих областях биомедицинской науки, являются базы данных. Пополнение баз данных, как правило, производится вручную путем экспертной оценки первоисточников информации – текстов научных публикаций. Алгоритмы интеллектуального анализа текстов позволяют существенно уменьшить финансовые и временные затраты для пополнения баз данных по биомедицинской информации. В большинстве алгоритмов интеллектуального анализа текстов можно выделить следующие этапы: (1) отбор релевантных текстов (статьи, патенты), (2) распознавание наименований целевых объектов (биологических и/или химических), (3) поиск ассоциаций между распознанными объектами, и (4) обработка и представление извлеченной информации для дальнейшего использования. Большинство из разработанных и свободно доступных методов реализуют отдельные из перечисленных этапов. Немногие из интегральных алгоритмов зачастую фокусируются на решении конкретной проблемы (например, молекулярных механизмах [1] или репозиционировании лекарственных препаратов для терапии болезни Паркинсона [2]). Поэтому целью данной работы стала разработка подхода, позволяющего извлекать всесторонние сведения о биологической активности химических соединений из текстов научных публикаций, и его апробация на примере противовирусных препаратов.

Методы и алгоритмы: Реализация разрабатываемого метода произведена на скриптовом языке Python (версия 3.11) в виде отдельных функций. Отбор препаратов для дальнейшего поиска сведений о биологической активности проводился следующим образом: были сформированы автоматические запросы к базе данных ChEMBL на основе раздела Drug Indications, среди которых упоминается термин “viral diseases” – для противовирусных препаратов. Дополнительно из БД ChEMBL нами были извлечены идентификаторы, известные мишени и возможные синонимы найденных лекарственных препаратов. С целью поиска текстов для анализа использована база данных библиографической информации PubMed, а в качестве единицы текста – аннотации научных публикаций. В основу запроса в БД PubMed легли наименования лекарственных препаратов и их синонимы, а также MeSH-термины (1) “Animals” – для исследований *in vivo*, (2) “Cells” – для исследований *in vitro*. Дополнительно мы собрали коллекцию текстов, содержащих клинические данные; с этой целью была

использована фильтрация по типам публикаций. Литературные обзоры были исключены на этапе запросов. Распознавание наименований реализовано с применением ранее разработанного алгоритма на основе метода условных случайных полей (CRF) [3–6] и библиотеки flair [7], а также регулярных выражений и словарей. В результате нами были извлечены наименования следующих объектов: химические соединения, белки/гены, виды, клеточные линии, заболевания, миРНК и однонуклеотидные полиморфизмы. С целью извлечения ассоциаций между перечисленными объектами разработан список правил, которые основаны на использовании семантической связи в текстах, например, «inhibits», «regulate», «ameliorate». Мы выбрали типы ассоциаций между объектами, значимые, на наш взгляд, для понимания биологической активности лекарственных препаратов, такие как: (1) химическое соединение-химическое соединение (метаболизм, межлекарственные взаимодействия), (2) химическое соединение-белок (терапевтические мишени, изменение клинических показателей), (3) химическое соединение-заболевание (применение в терапии, лекарственно-индуцированные патологии), (4) белок-белок (молекулярные механизмы, комплексы биомаркеров заболеваний), (5) химическое соединение-вид (индуцирующее/ингибирующее воздействие на микроорганизм, экспериментальные условия тестирования), (6) заболевание-заболевание (сопутствующие патологические состояния). С целью разграничения механизмов и эффектов, причин и следствий воздействия одного объекта на другой, семантические правила были разделены на категории и приведены к одному, наиболее общему понятию. К примеру, семантические связи «ameliorate», «alleviate», «correct» для ассоциации «химическое соединение-заболевание» отражают некоторое положительное воздействие лекарственного препарата на патогенез. Для исключения дубликатов объектов (например, вследствие, использования синонимов) реализованы запросы по наименованиям их объектов к фактографическим базам данных и онтологиям (ChEMBL [8], UniProt [9], PubChem [10], NCBI Taxonomy [11], Human Disease Ontology [12]). Извлеченные ассоциации были представлены в табличном виде, где каждая ассоциация имеет формат: *<объект 1><обобщенная семантическая связка><объект 2>* (напр., «imidapril inhibit angiotensin-converting enzyme»). На основании полученных таблиц сформированы файлы, формат которых приемлем для графического представления сетей взаимосвязей с применением платформы CytoScape [13].

Результаты: После запросов к ChEMBL с термином «viral diseases» среди назначений препаратов было отобрано 111 низкомолекулярных соединений. Среди противовирусных агентов мы отобрали только соединения с прямым противовирусным действием, против ВИЧ, вирусов гепатита С и В, SARS-CoV 2, гриппа и простого герпеса. Для указанных вирусов в БД ChEMBL содержится более 60 % соединений с противовирусной активностью. Итого, после фильтрации по организму, в список для анализа вошли 77 противовирусных препаратов. Запросы к PubMed привели к значительным пересечениям текстов по отдельным противовирусным препаратам, что неудивительно, поскольку, зачастую, они назначаются в комбинации. После фильтрации повторяющихся текстов, в коллекцию вошли более 150 тысяч аннотаций, релевантных изучению биологической активности противовирусных препаратов. Средняя точность распознавания наименований биологических и химических объектов с применением комбинации методов достигла 0.87. Для поиска ассоциаций нами

сформулировано более 500 правил, которые позволяют идентифицировать взаимосвязи «объект–активность–объект» со средней точностью, равной 0.78. Сети извлеченных взаимосвязей визуализированы с применением программы CytoScape. При ретроспективной верификации результатов обнаружены как сведения о препаратах, которые были тщательно изучены ранее, так и менее представленные в литературе материалы. В частности, найдены результаты исследований, проводимых в течение нескольких лет, о взаимосвязи длительного приема абакавира и проявлений атеросклероза и ремоделирования стенки артерий у молодых людей, живущих с ВИЧ-инфекцией [14–16]. Изучаются возможности репозиционирования противогрибкового препарата позаконазола в качестве противовирусного средства [17]. Дополнительно мы обнаружили, что в комбинации с ганцикловиром наблюдается синергетический эффект в подавлении цитомегаловирусной инфекции [17]. Для цидофовира – препарата, основным показанием к применению которого являются заболевания, вызванные цитомегаловирусом – при сочетанном применении с антидепрессантами - флувоксамином, венлафаксином, дулоксетином и бупропионом, и нормотимическими средствами – препаратами лития, карбамазепином, ламотриджином, показано влияние на замедление процесса их экскреции [18]. Была продемонстрирована активность адефовира (адефовир дипивоксил) в отношении клеточной линии OCI/AML-2 острого миелоидного лейкоза, а в комбинации с цитарабином (цитостатический препарат) исследователи наблюдали усиление эффекта [19].

Выводы: Разработанный интегральный подход применен к извлечению знаний о биологической активности низкомолекулярных соединений на примере противовирусных препаратов. Ретроспективная верификация результатов по данным литературы показала, что подход позволяет быстро и эффективно извлекать как широко изученные, так и менее известные эффекты и механизмы лекарственных препаратов.

Финансирование: Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 гг.) (№ 124050800018-9).

Application of text mining methods to extract comprehensive information about the biological activity of drugs: case-study for antiviral compounds

Biziukova N.*, Sobolev B., Karasev D., Ionov N., Sukhachev V., Taktashov R., Rudik A., Ivanov S., Tarasova O.

Institute of Biomedical Chemistry, Moscow, Russia

** nad.smol@gmail.com*

Key words: biological activity; text-mining; antiviral drugs; antihypertensive drugs; named entity recognition; relation extraction

Motivation and Aim: Databases are the main source of knowledge required in many areas of biomedical science. As a rule, databases are replenished manually by expert evaluation of information primary sources – scientific publications. Text mining

algorithms allow to significantly reduce time costs for replenishment of biomedical information databases. In the majority of text-mining algorithms the following stages can be distinguished: (1) selection of relevant texts (articles, patents), (2) named entity recognition (biological and/or chemical objects), (3) search for associations between the recognized entities, and (4) processing and representation of the extracted information for further use. Most of the developed and freely available methods are implementations of individual of the above listed steps; few of the integral algorithms often focus on solving a specific problem (e.g., molecular mechanisms [1], repositioning drugs for Parkinson's disease therapy [2]). Therefore, the aim of this work was to develop an approach that allows to extract comprehensive information about the biological activity of chemical compounds from the texts of scientific publications, and its validation within case-study for antiviral drugs.

Methods and Algorithms: The developed method was implemented using Python (version 3.11) in the form of separate functions. The selection of drugs for further search of information on biological activity was performed as follows: we generated automatic queries to the ChEMBL database based on the Drug Indications section, among which the term “viral diseases” was mentioned. Additionally, we extracted identifiers, known targets and possible synonyms of the found drugs from the ChEMBL database. For the purpose of text retrieval, the PubMed bibliographic information database was used for analysis, and abstracts of scientific publications were chosen as the text unit. The PubMed database query was based on drug names and their synonyms, as well as MeSH terms (1) “Animals” for in vivo studies, (2) “Cells” for in vitro studies. Additionally, we assembled a collection of texts containing clinical data; for this purpose, filtering by publication type was used. Literature reviews were excluded with the query. Named entity recognition was implemented using a previously developed algorithm based on the conditional random field (CRF) method [3–6] and the flair library [7], as well as regular expressions and dictionaries. As a result, we extracted the names of the following entities: chemical compounds, proteins/genes, species, cell lines, diseases, miRNAs and single nucleotide polymorphisms. In order to extract associations between the listed objects, we developed a set of rules that are based on the use of semantic links in texts (e.g., “inhibits”, “regulate”, “ameliorate”). We selected the types of associations between objects that we believe are significant for understanding the biological activity of drugs such as: (1) chemical compound-chemical compound (metabolism, drug-drug interactions), (2) chemical compound-protein (therapeutic targets, changes in clinical parameters), (3) chemical compound-disease (therapeutic applications, drug-induced pathologies), (4) protein-protein (molecular mechanisms, disease biomarker complexes), (5) chemical compound-species (inducing/inhibitory effects on microorganism, experimental testing conditions), (6) disease-disease (associated pathologic conditions). In order to distinguish between mechanisms and effects, causes and consequences of the influence of one object on another, semantic rules were categorized and brought to one, the most general concept. For example, the semantic relations “ameliorate”, “alleviate”, “correct” for the association “chemical compound-disease” reflect some positive effect of the drug on pathogenesis. To avoid duplicate entities (e.g., due to, for example, the use of synonyms), queries on the names of their entities to databases and ontologies (ChEMBL [8], UniProt [9], PubChem [10], NCBI Taxonomy 11], Human Disease Ontology [12]) were implemented. The extracted associations were tabulated, where each association has the format: *<object 1><generalized semantic conjunction><object 2>* (e.g., “imidapril inhibit angiotensin-

converting enzyme"). Based on the obtained tables, files were generated, the format of which is acceptable for graphical representation of interaction networks using the CytoScape platform [13].

Results: After queries to ChEMBL for the designation "viral diseases", 111 low molecular weight compounds were selected. Among antiviral agents, we selected only compounds with direct antiviral activity against HIV, hepatitis C and B, SARS-CoV-2, influenza and herpes simplex. For the above viruses, the ChEMBL database contained more than 60 % of compounds with antiviral activity. In total, after filtering by organism, 77 antiviral drugs were included in the list for analysis. Queries to PubMed resulted in significant overlap of texts for both individual antiviral drugs, which is not surprising since, often, they are prescribed in combination. After filtering out repetitive texts, the collection included more than 150,000 abstracts relevant to the study of the biological activity of antiviral drugs. The average accuracy of named entity recognition for chemical and biological objects using the combination of methods reached 0.87. To retrieve associations, we formulated more than 500 rules that allow us to identify object–acti-vity–object chains with an average accuracy of 0.78. The extracted association networks were visualized using CytoScape software. Retrospective verification of the results revealed both information on drugs that had been thoroughly studied earlier and less represented in the literature. In particular, we found the results of studies conducted over several years on the relationship between long-term administration of abacavir and manifestations of atherosclerosis and arterial wall remodeling in young people living with HIV infection [14–16]. The repositioning of the antifungal drug posaconazole as an antiviral agent is being investigated [17]. Additionally, we found that in combination with ganciclovir, a synergistic effect was observed in suppressing cytomegalovirus infection [17]. For cidofovir, a drug whose main indication for use is diseases caused by cytomegalovirus, when used in combination with antidepressants – fluvoxamine, venlafaxine, duloxetine and bupropion, and normotimics – lithium, carbamazepine, lamotrigine, an effect on slowing down their excretion was shown [18]. The activity of adefovir (adefovir dipivoxil) was demonstrated against the OCI/AML-2 cell line of acute myeloid leukemia, and in combination with cytarabine (a cytostatic drug), researchers observed an enhanced effect [19].

Conclusion: The developed integral approach was applied to the extraction of knowledge about the biological activity of low molecular weight compounds using antiviral drugs as an example. Retrospective verification of the results against literature data showed that the approach allows for fast and efficient extraction of both widely studied and less known effects and mechanisms of drugs.

Funding: The work was performed within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030) (No.124050800018-9).

Список литературы/References

1. Bachman J.A., Gyori B.M., Sorger P.K. Automated assembly of molecular mechanisms at scale from text mining and curated databases. *Mol Syst Biol.* 2023;19(5):e11325. doi 10.15252/msb.202211325
2. Tandra G. et al. Literature-Based Discovery Predicts Antihistamines Are a Promising Repurposed Adjuvant Therapy for Parkinson's Disease. *Int J Mol Sci.* 2023;24(15):12339. doi 10.3390/ijms241512339
3. Tarasova O. et al. Identification of Molecular Mechanisms Involved in Viral Infection Progression Based on Text Mining: Case Study for HIV Infection. *Int J Mol Sci.* 2023;24(2):1465. doi 10.3390/ijms24021465
4. Tarasova O. et al. Extraction of Data on Parent Compounds and Their Metabolites from Texts of Scientific Abstracts. *J Chem Inf Model.* 2021;61(4):1683-1690. doi 10.1021/acs.jcim.0c01054

5. Biziukova N. et al. Automated Extraction of Information From Texts of Scientific Publications: Insights Into HIV Treatment Strategies. *Front Genet.* 2020;11:618862. doi 10.3389/fgene.2020.618862
6. Biziukova N., Ivanov S., Tarasova O. Identification of Proteins and Genes Associated with Hedgehog Signaling Pathway Involved in Neoplasm Formation Using Text-Mining Approach. *Big Data Mining Analytics*. 2024;7(1):107-130. doi 10.26599/BDMA.2023.9020007
7. Weber L. et al. HunFlair: an easy-to-use tool for state-of-the-art biomedical named entity recognition. *Bioinformatics*. 2021;37(17):2792-2794. doi 10.1093/bioinformatics/btab042
8. Davies M. et al. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.* 2015;43(W1):W612-W620. doi 10.1093/nar/gkv352
9. Apweiler R. et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 2004;32:D115-D119. doi 10.1093/nar/gkh131
10. Kim S. et al. PubChem 2023 update. *Nucleic Acids Res.* 2023;51(D1):D1373-D1380. doi 10.1093/nar/gkac956
11. Federhen S. The NCBI Taxonomy database. *Nucleic Acids Res.* 2012;40:D136-D143. doi 10.1093/nar/gkr1178
12. Schriml L.M. et al. Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res.* 2012;40:D940-D946. doi 10.1093/nar/gkr972
13. Shannon P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-2504. doi 10.1101/gr.1239303
14. Dirajlal-Fargo S. et al. Longitudinal Changes in Subclinical Vascular Disease in Ugandan Youth With Human Immunodeficiency Virus. *Clin Infect Dis.* 2023;76(3):e599-e606. doi 10.1093/cid/ciac686
15. Pocock M.O. et al. Pathophysiology of ischaemic heart disease. *Curr Opin HIV AIDS*. 2017;12(6):548-553. doi 0.1097/COH.0000000000000411
16. Hsue P.Y. et al. Association of abacavir and impaired endothelial function in treated and suppressed HIV-infected patients. *AIDS.* 2009;23(15):2021-7. doi 10.1097/QAD.0b013e32832e7140
17. Mansuri Z. et al. Drug Interactions Between Psychotropic Medications and Treatment of Mpox (tecovirimat and cidofovir). *Prim Care Companion CNS Disord.* 2024;26(1):23lr03608. doi 10.4088/PCC.23lr03608
18. Mercorelli B. et al. The Clinically Approved Antifungal Drug Posaconazole Inhibits Human Cytomegalovirus Replication. *Antimicrob Agents Chemother*. 2020;64(10):e00056-20. doi 10.1128/AAC.00056-20
19. Khoury H. et al. Octadecyloxyethyl Adefovir Exhibits Potent *in vitro* and *in vivo* Cytotoxic Activity and Has Synergistic Effects with Ara-C in Acute Myeloid Leukemia. *Chemotherapy*. 2018;63(4):225-237. doi 10.1159/000491705

doi 10.18699/bgrs2024-3.2-02

Молекулярные взаимодействия пептидов β-амилоидов, как неупорядоченных белков, и перспективных лекарственных средств на основе D-энантиомерных пептидов

Бочаров Э.В.[1, 2]*, Охрименко И.С.[1], Волынский П.Е.[2], Павлов К.В.[1], Злобина В.В.[1], Бершацкий Я.В.[1, 2], Крючкова А.К.[1, 2], Кузьмичев П.К.[1], Ефремов Р.Г.[1, 2]

[1] *Московский физико-технический институт (национальный исследовательский университет), Долгопрудный, Россия*

[2] *Институт биоорганической химии им. ак. М.М. Шемякина и Ю.А. Овчинникова РАН, Москва, Россия*

* *edvbon@mail.ru*

Ключевые слова: терапия болезни Альцгеймера; белок-предшественник амилоида; трансмембранный домен; белок-липидные и белок-белковые взаимодействия; ЯМР-спектроскопия; молекулярная динамика

Мотивация и цель: Болезнь Альцгеймера (БА) – разрушительное нейродегенеративное заболевание, приводящее к тяжелой деменции. Детальная информация о структуре, динамике и межмолекулярных взаимодействиях биомолекул, непосредственно участвующих в развитии болезни Альцгеймера, необходима для рациональной разработки новых биологически активных соединений и скрининга существующих с целью получения наиболее эффективных кандидатов в лекарственные средства [1]. Генетические данные убедительно свидетельствуют о том, что аберрантное образование, агрегация и/или клиренс нейротоксических пептидов амилоида-β (Aβ), являющегося продуктом последовательного расщепления белка-предшественника амилоида (APP), запускает заболевание. Aβ аккумулируется в местах контакта нейронов в токсические олигомеры и фибриллы, образуя так называемые сенильные бляшки. При этом изоформы Aβ различной длины обнаруживаются в мозгу здоровых людей независимо от возраста и, по-видимому, играют роль в сигнальных путях в головном мозге и обладают нейропротекторными свойствами при низких концентрациях. D-энантиомерный пептид D3 и его производные недавно были отобраны коллегами из университета г. Дюссельдорф (Германия) с помощью фагового дисплея для прямого разрушения цитотоксических агрегатов Aβ [2]. В настоящее время одно из D3-подобных соединений должно пройти фазу II клинических испытаний, однако молекулярные детали его профилактического или фармакологического действия не совсем ясны.

Методы и алгоритмы: Для решения проблемы мы использовали комплексный подход, основанный на биохимических и биофизических методах, таких как белковая инженерия, микроскопический термофорез, флуоресцентная конфокальная микроскопия, флуоресцентная поляризация, калибровка микрофлюидной диффузии, круговой дихроизм, спектроскопия ядерного магнитного резонанса высокого разрешения (ЯМР) и компьютерное моделирование.

Результаты: В работе мы представляем экспериментальные данные, свидетельствующие о том, что D3-пептид, будучи внутренне неупорядоченным пептидом (IDP), может динамически и специфически связываться IDP/IDP-подобным образом с внеклеточной частью мембраносвязанного Aβ-предшественника, трансмембранным CTFβ фрагментом белка-предшественника амилоида ($APP_{672-726}$, $Aβ_{1-55}$ в нумерации β-амилоида). В частности, гетероядерная ЯМР-спектроскопия показала, что D3-пептид напрямую связывается с амфифильной примембранным JM-участком $Aβ_{17-26}$, который в зависимости от внешних условий способен претерпевать конформационный переход из α-спиральной в β-конформацию и участвует в свертывании β-амилоида в фибриллы. Также было протестировано влияние ряда патогенных мутаций БА, локализованных в различных структурных и функциональных частях APP, на связывание D3-пептида.
Выводы: Полученные данные свидетельствуют о том, что D-энантиомерный пептид D3 распознает амилоидогенный участок мембранного белка APP и его фрагментов – Aβ пептидов, ограничивая их конформационное разнообразие, не нарушая α-спиральности и предотвращая образование межмолекулярных водородных связей. Это создает предпосылки для ингибирования ранних стадий превращения Aβ пептидов в β-конформацию и последующей токсической олигомеризации, связанной с ранними стадиями развития болезни Альцгеймера [3]. Достигнутый прогресс в понимании молекулярного механизма действия D3-пептида является важным шагом на пути к разработке эффективной стратегии лечения и профилактики болезни Альцгеймера.
Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 23-74-00024, https://rscf.ru/project/23-74-00024/.

Molecular interactions of β-amyloid peptides, as disordered proteins, and promising drugs based on all-D-enantiomeric peptides

Bocharov E.V.[1, 2]*, Okhrimenko I.S.[1], Volynsky P.E.[2], Pavlov K.V.[1], Zlobina V.V.[1, 2], Bershatsky Ya.V.[1, 2], Kryuchkova A.K.[1], Kuzmichev P.K.[1], Efremov R.G.[1, 2]

[1] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

[2] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

* *edvbon@mail.ru*

Key words: Alzheimer´s disease treatment; amyloid precursor protein; transmembrane domain; protein-lipid and protein-protein interactions; NMR spectroscopy; molecular dynamics

Motivation and Aim: Alzheimer's disease (AD) is a devastating neurodegenerative disease that leads to severe dementia. Detailed information on the structure, dynamics and intermolecular interactions of biomolecules directly involved in the development of Alzheimer's disease is necessary for the rational development of new biologically active compounds and screening of existing ones in order to obtain the most effective drug candidates [1]. Genetic evidence strongly suggests that aberrant production, aggregation, and/or clearance of neurotoxic amyloid-β (Aβ) peptides, being the products of sequential cleavage of transmembrane amyloid precursor protein (APP), triggers disease. Aβ

accumulates at the toxic oligomers and fibrils, forming so-called senile plaques. However, Aβ isoforms of varying lengths are found in the brain of healthy people regardless of age and appear to play a role in signaling pathways in the brain and have neuroprotective properties at low concentrations. All-D-enantiomeric peptide D3 and its derivatives were recently selected by the colleagues from the University of Dusseldorf (Germany) using phage display for direct destruction of cytotoxic Aβ aggregates [2]. Currently, one of the D3-like compounds is about to enter phase II clinical trials, but the molecular details of its prophylactic or pharmacological effects are not entirely clear.

Methods and Algorithms: To solve the problem, we used complex approach based on biochemical and biophysical methods such as protein engineering, microscopic thermophoresis, fluorescence confocal microscopy, fluorescence polarization, microfluidic diffusion calibration, circular dichroism, high resolution nuclear magnetic resonance spectroscopy (NMR) and computer simulation.

Results: We present experimental evidence showing that D3-peptide, being an intrinsically disordered peptide (IDP), can dynamically and specifically bind in IDP/IDP-like manner to the extracellular juxtamembrane (JM) region of membrane-bound Aβ precursor, CTFβ transmembrane fragment of amyloid precursor protein ($APP_{672-726}$, $Aβ_{1-55}$ in amyloid-β numeration). Namely, heteronuclear NMR spectroscopy showed that D3-peptide directly binds to the amphiphilic near-membrane JM region of $Aβ_{17-26}$, which, depending on external conditions, is capable of undergoing a conformational transition from the α-helical conformation to the β-chain, which is involved in folding of β-amyloid into fibrils. The influence of a number of pathogenic AD mutations located in different structural and functional parts of APP on the peptide binding was also tested.

Conclusion: The data suggest that D-enantiomeric peptide D3 recognizes the amyloidogenic region of APP and its fragments – Aβ peptides, restricting conformational diversity not compromising its α-helicity and preventing intermolecular hydrogen bond formation, which would create prerequisites for inhibition of early steps of Aβ conversion into β-conformation and its toxic oligomerization associated with early stages of AD development [3]. The achieved progress in understanding the molecular mechanism of D3-peptide action is an important step towards development of an effective AD treatment and prevention strategy.

Funding: The study was supported by the Russian Science Foundation grant No. 23-74-00024, https://rscf.ru/project/23-74-00024/.

Список литературы/References

1. Urban A.S. et al. Structural Studies Providing Insights into Production and Conformational Behavior of Amyloid-β Peptide Associated with Alzheimer's Disease Development. *Molecules*. 2021;26(10):2897
2. Klein A.N., Gremer L., Kutzsche J., Willbold D. et al. Optimization of d-Peptides for Aβ Monomer Binding Specificity Enhances Their Potential to Eliminate Toxic Aβ Oligomers. *ACS Chemical Neurosci*. 2017;8(9):1889-1900
3. Bocharov E.V. et al. All-d-Enantiomeric Peptide D3 Designed for Alzheimer's Disease Treatment Dynamically Interacts with Membrane-Bound Amyloid-β Precursors. *J Med Chem.* 2021;64(22):16464-16479

doi 10.18699/bgrs2024-3.2-03

Рациональные подходы создания высокоизбирательного и эффективного ингибитора главной протеазы SARS-CoV-2

Булыгин А.[1]*, Кузнецов Н.[1, 2]

[1] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

[2] *Факультет естественных наук, Новосибирский государственный университет, Новосибирск, Россия*

* *abulygin@niboch.nsc.ru*

Ключевые слова: SARS-CoV-2; главная протеаза; молекулярное моделирование; ингибитор

Мотивация и цель: Пандемия COVID-19, вызванная коронавирусом SARS-CoV-2, стала третьим случаем коронавирусной инфекции. Вместе с другими двумя вирусами, SARS-CoV и MERS-CoV, вирус SARS-CoV-2 является высокопатогенным и имеет уровень смертности более 1 %. Более того, стало понятным, что коронавирус часто мутирует, что снижает эффективность доступных вакцин и вынуждает регулярно создавать новые.

Вирусные полимераза и протеазы являются подходящими мишенями для лекарств прямого действия. На данный момент существует три одобренных противокоронавирусных лекарства: ремдесивир [1], молнупиравир [2] и нирматрелвир [3]. Однако оба лекарства, направленных против полимеразы, не рекомендованы для использования из-за незначительной эффективности в случае с ремдесивиром [4] и потенциального увеличения частоты мутаций вируса в случае с молнупиравиром [2]. Нирматрелвир, ингибитор главной протеазы (M^{pro}), стал более успешным, его начали выпускать в форме таблеток. Он имеет хорошую биодоступность – доля усваиваемого препарата равна примерно 50 %, но растворимость соединения находится на умеренном уровне (около 1 мг/мл) [3]. Таким образом, по-прежнему остается актуальным создание высокоизбирательного и эффективного ингибитора протеазы, который можно принимать перорально.

В данной работе мы провели глубокий литературный обзор *in vitro* и *in vivo* исследований ингибиторов M^{pro} и создали набор возможных ингибиторов и их частей, которые предположительно позволяют достичь всех необходимых свойств, а именно высокого сродства к вирусному ферменту, избирательности, биодоступности и растворимости.

Методы и алгоритмы: Молекулярно-динамическое моделирование было проведено в пакете программ GROMACS 2020.6 в сочетании с силовым полем AMBER ff99SB-ILDN [5]. Свободные энергии связывания комплексов были вычислены сочетанием методов молекулярной механики и Пуассон-Больцмановской площади поверхности. Для вычисления этим методом был использован скрипт “g_mmpbsa” для формата файлов GROMACS [6]. Онлайн-сервис SwissADME был использован для расчета лекарственных характеристик: всасываемости, распределения, метаболизма и выделения [7].

Результаты: Было проведено МД-моделирование комплексов основной протеазы коронавируса с несколькими известными ингибиторами с различным уровнем эффективности: PF-0835231 [8], нирматрелвир, MPI3 [9], GC373 [10], боцепревир и телапревир [11]. Анализ образуемых водородных связей между данными

ингибиторами и аминокислотными остатками активного центра M^{pro} показал, что наибольшее среднее по времени значение водородных связей имеет PF-0835231 – 6.6 связи, тогда как максимально возможное для него значение равно 9. Энергия связывания PF-0835231 в активном центре фермента равна –141 ± 13 кДж/моль. Благодаря глубокому проникновению в связывающий центр фермента, наилучшую энергию связывания (–189 ± 14 кДж/моль) имеет соединение MPI3. Однако IC_{50} для MPI3 (8 нМ) хуже, чем для PF-0835231 (0.3 нМ), что может объясняться высокой гибкостью молекулы MPI3, снижающей эффективность образования комплекса с ферментом. Недостатком высокоэффективного ингибитора PF-0835231 является его неудовлетворительная растворимость (менее 0.1 мг/мл), которая может быть повышена фосфорилированием молекулы. Однако такая модификация приводит к практически полной потере биодоступности [3]. Согласно полученным данным, нирматрелвир является третьим по энергии связывания (–138 ± 13 кДж/моль), имея IC_{50} = 3 нМ. Таким образом, можно заключить, что свободная энергия связывания соединений в активном центре фермента на уровне –140 кДж/моль и ниже может обеспечить эффективное ингибирование протеазы, однако индивидуальные структурные особенности ингибитора могут иметь существенное значение для создания эффективного и селективного препарата.

Анализ полученных данных позволяет предположить, что ингибитор M^{pro} должен обладать пятью главными качествами, придающими высокий уровень связывания и избирательности (рис. 1): 1) заместитель S0 должен быть небольшим; 2) заместитель S1 должен быть похож на глутамин, имея одного донора и одного акцептора водородной связи; 3) боковая цепь S2 должна быть гидрофобной и иметь длину в три-четыре C-C связи; 4) для повышения избирательности по отношению к M^{pro} участок S2 должен иметь структуру боковой цепи, жестко связанную с остовом (например, бициклопролин), или должен быть жестким и объемным заместитель S4 (например, индол, содержащий заместитель в четвертом положении); 5) заместитель S4 должен повторять форму кармана S4 связывающего центра. При этом значительных ограничений структуры заместителя в положении S3 не было обнаружено.

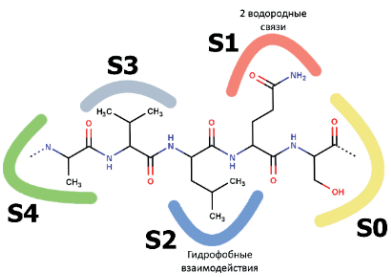


Рис. 1. Общая структура субстрата M^{pro}

Формулирование таких критериев для эффективного ингибитора позволяет пойти двумя путями улучшения свойств предложенных ранее соединений. Один из них заключается в создании аналогов PF-0835231, обладающих хорошей растворимостью без потери эффективности связывания. Для этого в ходе данной работы были проанализированы аналоги PF-0835231, несущие различные модификации в положениях S0, S1 и S4, затем для всех модифицированных соединений рассчитывали растворимость с помощью SwissADME и наконец проводили моделирование комплекса созданных ингибиторов с ферментом. Лучший результат показало соединение, в котором в качестве S4-заместителя был фторированный индол, а в качестве S1 – остаток сукцинимида (рис. 2). Кроме того, гидроксиметилкетоновая группа, отвечающая за образование ковалентного комплекса с ферментом, была заменена на альдегидную (на рис. 2 выделена

кружком), так как в литературе было показано, что альдегидный вариант обладает большей биодоступностью [12]. Итоговое соединение имело энергию связывания –156 ± 13 кДж/моль, что на 15 кДж/моль лучше значения для PF-0835231. При этом количество образуемых водородных связей в активном центре не изменилось относительно PF-0835231, расчетная растворимость улучшилась примерно в два раза (с 0.1 до 0.2 мг/мл), а биодоступность стала высокой согласно расчету SwissADME. Данные характеристики позволяют предположить, что предложенное соединение может обладать улучшенной эффективностью действия в сравнении с исходным ингибитором PF-0835231. Еще одним способом улучшения растворимости любого рассмотренного ингибитора оказалась замена пирролидона в положении S1 на имидазолидон (см. рис. 2). Согласно SwissADME, это изменение ведет к улучшению растворимости в 3–5 раз в зависимости от алгоритма расчета растворимости. Данная модификация большинства рассмотренных соединений, по результатам моделирования, не приводит к изменениям ни в положении ингибиторов в активном центре, ни в количестве образуемых водородных связей, ни в величине энергии связывания.

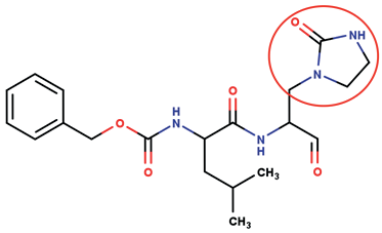


Рис. 2. Пример структуры соединения с имидазолидоном (выделен кружком) в качестве S1

Второй путь улучшения известных ингибиторов заключается в локальном изменении структуры ингибитора для увеличения эффективности связывания с сохранением хорошей растворимости. Перспективной основой для работы в данном направлении может выступить молекула нирматрелвира, поскольку ее растворимость уже находится на приемлемом уровне (1 мг/мл). Дальнейшее увеличение этого параметра может сопровождаться нежелательным снижением биодоступности, которая не может быть предсказана точно имеющимися в данный момент алгоритмами расчета. Поэтому мы решили попытаться улучшить связывание нирматрелвира с ферментом. Остаток бициклопролина, входящий в нирматрелвир и повышающий избирательность к M^{pro} (пункт 4 критериев эффективности), открывает возможность использования широкого набора S4-заместителей. Наилучшие результаты в качестве заместителя в положении S4 показал остаток тиофена (рис. 3), приводя к понижению свободной энергии связывания на 42 кДж/моль, до –180 ± 15 кДж/моль, в то время как растворимость модифицированного соединения согласно расчету SwissADME практически не изменилась.

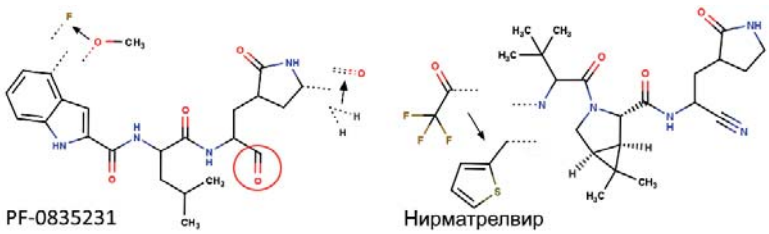


Рис. 3. Структуры PF-0835231 и нирматрелвира с обозначением наилучших модификаций

Вывод: В данной работе методами молекулярного моделирования проведен структурно-функциональный анализ ингибиторов главной протеазы коронавируса и предложены способы улучшения растворимости и эффективности связывания соединений на основе структурно-модифицированных аналогов известных ингибиторов.
Финансирование: Исследование выполнено в рамках государственного задания ИХБФМ СО РАН (№ 121112900214-2).

Approaches to rational design of a highly selective and potent SARS-CoV-2 main protease inhibitor

Bulygin A.[1]*, Kuznetsov N.[1, 2]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

* *abulygin@niboch.nsc.ru*

Key words: SARS-CoV-2; main protease; inhibitor; molecular modelling; rational design

Motivation and Aim: COVID-19 caused by the SARS-CoV-2 is the third case of coronavirus pandemics. Along with the other two, SARS-CoV and MERS-CoV, it is highly pathogenic and has mortality rates more than 1 %. In addition, it became apparent that coronavirus has high mutability that decrease the effectiveness of available vaccines and makes necessary to regularly develop new ones.
The viral polymerase and proteases are applicable as direct action targets of anticoronavirus drugs. Currently there are three approved anti-coronavirus drugs: remdesivir [1], molnupiravir [2] and nirmatrelvir [3]. However, both polymerase-directed drugs are not recommended for wide use because of negligible efficacy in the case of remdesivir [4] or potentially increased mutability of the virus in the case of molnupiravir [2]. Nirmatrelvir, the inhibitor of the main protease (M^{pro}), has become more successful being produced in tablet form. It has good bioavailability (about 50 %) but only moderate solubility under 1 mg/ml [3]. Thus, there is still a need in a highly selective and potent M^{pro} inhibitors, that can be administered orally.
In this work we conducted a deep review of *in vitro* and *in vivo* studies of M^{pro} inhibitors and designed a set of parts of possible inhibitors, which presumably achieved all necessary properties, namely binding ability, selectivity, bioavailability and solubility.
Methods and Algorithms: Molecular dynamics simulations were performed with GROMACS 2020.6 and the AMBER ff99SB-ILDN force field [5]. Binding free energies of complexes were calculated by combining MD data with the molecular mechanic/Poisson-Boltzmann surface area method. The "g_mmpbsa" script for GROMACS data [6] was used to calculate the free energies of MM-PBSA. SwissADME online tool was used for prediction of ADME properties [7].
Results: MD simulations of several known inhibitors with different inhibition efficiency were conducted. The chosen inhibitors were: PF-0835231 [8], nirmatrelvir, MPI3 [9], GC373 [10], boceprevir and telaprevir [11]. The analysis of the H-bonds formed between the inhibitors and M^{pro} showed that the highest time average number of H-bonds has PF-0835231 – 6.6 bonds with a maximum of 9. MPI3 has the best binding energy (–189 ± 14 kJ/mol) thanks to penetration deep into the active site, PF-0835231 is on the second place (–141 ± 13 kJ/mol). However, MPI3 has worse IC_{50} than that of PF-

0835231 (8 nM against 0.3 nM), that may be explained by the high flexibility with a lot of rotational degrees of freedom that is considered as a bad factor for binding. On the other hand, PF-0835231 has an inappropriate solubility (less than 0.1 mg/ml), that can be increased by phosphorylation, which leads to an extremely low oral bioavailability [3]. Nirmatrelvir takes third place in binding with binding energy of –138 ± 13 kJ/mol, while its IC_{50} = 3 nM.

Taking together all the knowledge it could be stated that M^{pro} inhibitor should have five main properties (Fig. 1): (1) small S0-residue; (2) glutamine-like S1-residue with one H-bond acceptor and one H-bond donor; (3) S2-residue with the length of 3 or 4 C-C bonds; (4) a bulky part of a backbone (like a bicycloproline) or bulky and rigid S4-residue (like an indole with substituents) for selectivity; (5) S4-residue repeating the shape of the enzyme's S4-pocket for binding.

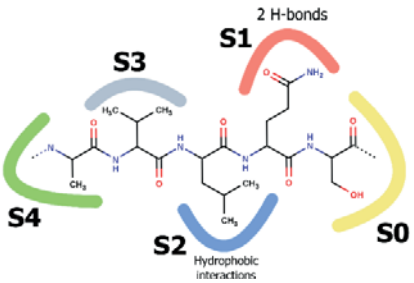


Fig. 1. The general structure of M^{pro} substrate

Formulating these criteria allows us to go two ways to improve known inhibitors. First, we tried to improve the solubility of PF-0835231 without decrease in binding. The best result showed a molecule, in which we used fluoride indole as S4-substituent and succinimide as S1 (Fig. 2). We also changed S0-substituent intended for covalent binding ("warhead") from hydroxymethylketone to an aldehyde because it was shown that the inhibitors with aldehyde warhead have better bioavailability [12]. The final compound has binding energy of –156 ± 13 kJ/mol that is by 15 kJ/mol less than that of PF-0835231. Number of H-bonds didn't change and SwissADME service showed that these changes improved the solubility twofold (from approximately 0.1 to 0.2 mg/ml) and bioavailability became high.

Another way to improve the solubility of any inhibitor is to replace a pyrrolidone S1-moiety by an imidazolidone one (Fig. 2). According to SwissADME this change makes the solubility 3–5 times higher depending on the solubility calculation algorithm. We modeled most of the compounds with this modification. Results showed that there are no changes in position, binding energy or a number of H-bonds for all tested compounds.

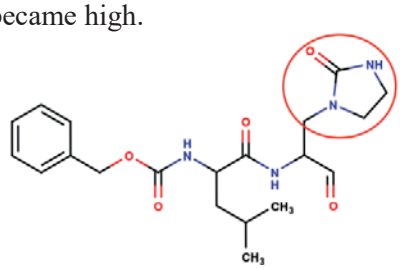


Fig. 2. Example structure of a compound with an imidazolidone S1-substituent

The second way to improve known inhibitors is to improve binding ability while maintaining the solubility on the same level. Nirmatrelvir is a good basis for this. This compound already has an acceptable solubility (about 1 mg/ml), but, on the other hand, further increase of this parameter may lead to bioavailability decrease, which cannot be well predicted yet. Thus, we made an attempt to improve binding of nirmatrelvir. Bicycloproline, which confers selectivity by itself, opens up the possibility of using much wider range of S4-substituents. Best results showed the compound with a small aromatic tiophene as S4 (Fig. 3). It has binding energy of –180 ± 15 kJ/mol that is 42 kJ/mol lower than that of nirmatrelvir, while its solubility almost didn't change.

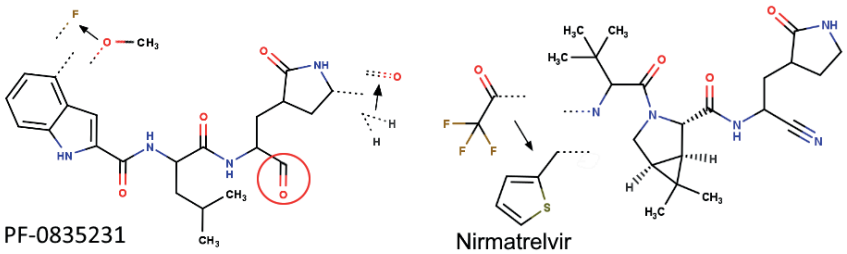


Fig. 3. Structures of PF-0835231 and nirmatrelvir with modifications that give the greatest improvement

Conclusions: Thus, in this work we conducted structure-function analysis of SARS-CoV-2 main protease inhibitors using molecular modeling methods and propose approaches of improving the solubility and binding ability of compounds based on structurally modified analogues of known inhibitors.

Funding: This study was supported by the Russian state-funded project for ICBFM SB RAS (No. 121112900214-2).

Список литературы/References

1. Siegel D., Hui H.C., Doerffler E., Clarke M.O. et al. Discovery and Synthesis of a Phosphoramidate Prodrug of a Pyrrolo[2,1- *f*] [Triazin-4-Amino] Adenine *C*-Nucleoside (GS-5734) for the Treatment of Ebola and Emerging Viruses. *J Med Chem.* 2017;60:1648-1661. doi 10.1021/acs.jmedchem.6b01594
2. Cox R.M., Wolf J.D., Plemper R.K. Therapeutically Administered Ribonucleoside Analogue MK-4482/EIDD-2801 Blocks SARS-CoV-2 Transmission in Ferrets. *Nat Microbiol.* 2021;6:11-18. doi 10.1038/s41564-020-00835-2
3. Owen D.R., Allerton C.M.N., Anderson A.S., Aschenbrenner L., Avery M., Berritt S., Boras B., Cardin R.D. et al. An Oral SARS-CoV-2 M^{pro} Inhibitor Clinical Candidate for the Treatment of COVID-19. *Science.* 2021;374:1586-1593. doi 10.1126/science.abl4784
4. Sheahan T.P., Sims A.C., Zhou S., Graham R.L., Pruijssers A.J., Agostini M.L., Leist S.R., Schäfer A., Dinnon K.H. et al. An Orally Bioavailable Broad-Spectrum Antiviral Inhibits SARS-CoV-2 in Human Airway Epithelial Cell Cultures and Multiple Coronaviruses in Mice. *Sci Transl Med.* 2020;12(541):eabb5883. doi 10.1126/scitranslmed.abb5883
5. Abraham M.J., Murtola T., Schulz R., Páll S., Smith J.C., Hess B., Lindah E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX.* 2015;1-2:19-25. doi 10.1016/j.softx.2015.06.001
6. Kumari R., Kumar R., Lynn A. G_mmpbsa – A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J Chem Inf Model.* 2014;54:1951-1962. doi 10.1021/ci500020m
7. Daina A., Michielin O., Zoete V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci Rep.* 2017;7:42717. doi 10.1038/srep42717
8. Hoffman R.L., Kania R.S., Brothers M.A., Davies J.F. et al. Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J Med Chem.* 2020;63:12725-12747. doi 10.1021/acs.jmedchem.0c01063
9. Yang K.S., Ma X.R., Ma Y., Alugubelli Y.R., Scott D.A., Vatansever E.C., Drelich A.K., Sankaran B. et al. A Quick Route to Multiple Highly Potent SARS-CoV-2 Main Protease Inhibitors**. *Chem Med Chem.* 2021;16:942-948. doi 10.1002/cmdc.202000924
10. Vuong W., Fischer C., Khan M.B., van Belkum M.J., Lamer T., Willoughby K.D., Lu J., Arutyunova E. et al. Improved SARS-CoV-2 Mpro Inhibitors Based on Feline Antiviral Drug GC376: Structural Enhancements, Increased Solubility, and Micellar Studies. *Europ J Med Chem.* 2021;222:113584. doi 10.1016/j.ejmech.2021.113584
11. Anson B.J., Chapman M.E., Lendy E.K., Pshenychnyi S., D'Aquila R.T., Satchell K.J.F., Mesecar A.D. Broad-Spectrum Inhibition of Coronavirus Main and Papain-like Proteases by HCV Drugs. *Res Square.* 2020. doi 10.21203/rs.3.rs-26344/v1
12. Qiao J., Li Y.-S., Zeng R., Liu F.-L., Luo R.-H. et al. SARS-CoV-2 M^{pro} Inhibitors with Antiviral Activity in a Transgenic Mouse Model. *Science.* 2021;371:1374-1378. doi 10.1126/science.abf1611

doi 10.18699/bgrs2024-3.2-04

ГХ-МС исследование липофильных компонентов в листьях *Triticum aestivum* L.

Васильева А.Р.[1, 2]*, Слынько Н.М.[1, 2], Гончаров Н.П.[1], Татарова Л.Е.[1, 2], Куйбида Л.В.[3], Пельтек С.Е.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

[3] *Институт химической кинетики и горения им. В.В. Воеводского СО РАН, Новосибирск, Россия*

* *vasilieva@bionet.nsc.ru*

Ключевые слова: *Triticum aestivum*; ГХ-МС; полиморфизм; хемотаксономия; метаболизм

Мотивация и цель: Пшеница *Triticum aestivum* L. принадлежит к числу основных зерновых культур на земном шаре. Количество сортов и подвидов пшеницы огромно и продолжает увеличиваться. Метаболические исследования позволяют глубже понять фенотипическое разнообразие внутривидового полиморфизма *T. aestivum* и являются ключевыми в хемотаксономической классификации и идентификации видов и подвидов на основе их химического состава. Определение химического состава липофильных метаболитов в экстрактах зеленых листьев даст возможность установить взаимосвязи как внутри отдельных метаболических путей, так и между этими конкретными путями. Реализация такого сравнительного изучения на широкой выборке сортов и подвидов *T. aestivum* является целью данного исследования.

Методы и алгоритмы: Изучена коллекция 30 сортообразцов пшеницы из коллекции сектора генетики пшениц ИЦиГ СО РАН: гексаплоидной (мягкой *T. aestivum* и ее подвида *T. aestivum* ssp. *petropavlovskyi* N.P. Gontsch.), тетраплоидной (*T. cartlicum* Nevski № 21, 24) и диплоидной (*T. monococcum* L. № 26, 27). Для аннотирования соединений брались данные, полученные на системах: Pegasus 4D GCxGC–TOF MS (включает Agilent Technologies 7890B) и хроматограф Agilent Technologies 6890 с масс-спектрофотометрическим детектором 5973. Для визуализации различий между исследуемыми экстрактами листьев разных сортов пшеницы был использован иерархический кластерный анализ. Для более полного охвата и глубокого анализа изучаемых процессов применяли подходы многомерного анализа, в частности метод главных компонент (PCA).

Результаты: В результате интерпретации данных на основе анализа масс-спектров было определено, что наибольшим содержанием в большинстве экстрактов обладают алифатические спирты линейного строения и их производные. Углеводороды, как линейные, так и разветвленные, определены в диапазоне C10–C40, среди линейных преобладают C29, C31. Всего идентифицировано 16 насыщенных алифатических кислот, 5 их эфиров, 5 ненасыщенных кислот, 6 их эфиров, 2 лактона. Алифатические альдегиды и кетоны представлены 14 линейными насыщенными и 5 ненасыщенными альдегидами, 4 линейными и 2 кетонами с разветвленным строением, 4-Hydroxy-2-butanone, а также тремя β- и одним γ-дикетонами. Hentriacontane-14,16-dione и его гидроксипроизводное, согласно литературным данным, являются фактором осизости листьев пшеницы. Производные изопрена, обнаруженные на хроматограммах образцов, включают:

монотерпены (11) и образующиеся из них в растении цимолы (2), дитерпены (17), сесквитерпен (1), стероиды и их производные (13), а также их предшественники (2). К этой же группе относятся идентифицированные соединения, включающие в структуру полипреновый фрагмент: витамин Е, токоферолы α, γ, их эфиры, токоспиро А и Б, а также 4,8,12,16-tetramethylheptadecan-4-olide. Кроме того, на хроматограммах определены Geranylacetone, Farnesylacetone, Geranyllinalool. Ароматические соединения, идентифицированные в экстрактах: фенолы (13) и их производные (5), бензойная кислота и ее производные (3), ароматические системы, сопряженные с карбонильной группой (5) или с двойной связью (3), конденсированные системы (5). Кроме того, на хроматограммах обнаружен Benzeneacetaldehyde. Пики в масс-спектрах с индексами Ковача 2043, 2267, 2367, 2467, 2667, 2767, 2863 программой AMDIS, LECO ChromaTOF определяются либо как N-этиламиды, либо как N,N-диметиламиды тетрадекановой, гексадекановой, гептадекановой, октадекановой, эйкозановой, генэйкозановой и докозановой кислот соответственно. Дополнительные исследования утверждают структуру N-этиламидов жирных кислот.

Данные о метаболическом синтезе N-этиламидов жирных кислот отсутствуют. Можно предположить, что этот процесс включает в себя следующие стадии: 1 – взаимодействие соответствующей будущему амиду жирной кислоты с ацетил-КоА, протекающее с расходованием АТФ; 2 – реакция активированной кислоты с аланином с образованием N-ацилаланина жирной кислоты; 3 – реакция с участием фермента декарбоксилазы и пиридоксаль-5′-фосфата с образованием N-этиламида.

Расчет по методу PCA с использованием всех идентифицированых соединений позволил выделить несколько кластеров с возможным хемотаксономическим сходством (рис. 1). Образцы с преобладающим содержанием октакозанола можно выделить в отдельную группу, расположенную в зоне отрицательных значений PC1 (кластер 1), в отличие от образцов с преобладанием октакозанола и ацетатов фитола (кластер 2). Отдельно расположены образцы с преобладанием алкиламидов жирных кислот (кластер 3).

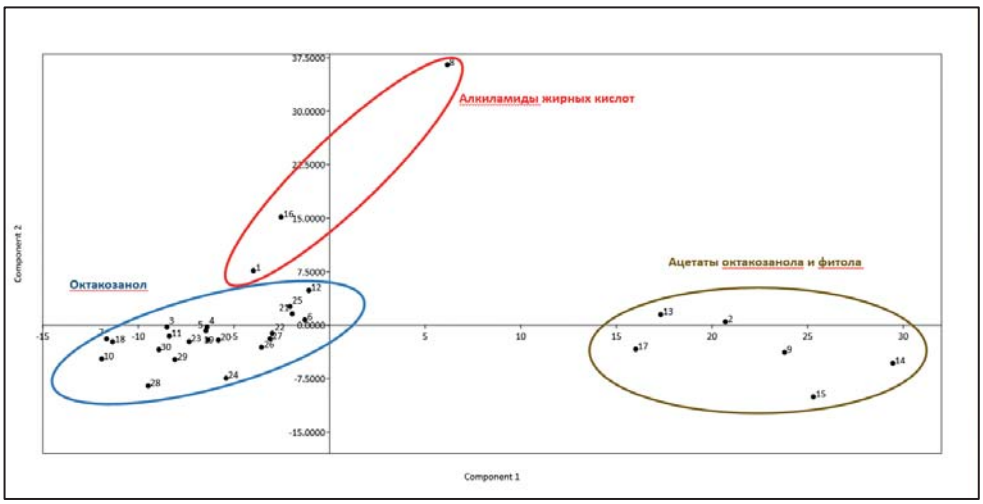


Рис. 1. Биплоты PCA на основе всех идентифицированных соединений в 30 сортах пшеницы. Внутренняя таблица содержит собственные значения и процент изменчивости для первых двух главных компонент

Ацилирование алифатических спиртов, преобладающих в кластере 1, в ацетаты (основные компоненты кластера 2) предположительно происходит в ответ на изменение влажности окружающей среды. Это ацилирование блокирует гидрофильные фрагменты алифатических спиртов, тем самым придавая кутикуле гидрофобные свойства. В результате эта модификация препятствует проникновению избыточной влаги в лист.

Выводы: В результате исследования составов липофильных фракций экстрактов листьев 30 сортов *Triticum aestivum* отмечено наличие хемотаксономической полиморфности.

Специфическими маркерами этих полиморфизмов являются: октакозанол, превалирующий в экстрактах большей части исследованных сортов, а также N-этиламиды жирных кислот, β-дикетоны и ацетаты октакозанола и фитола.

Впервые обнаруженные в большом количестве в части экстрактов образцов N-этиламиды жирных кислот могут играть важную роль в адаптации растений к абиотическим стрессам.

Корреляций между составом кутикулы и географическим происхождением не обнаружено.

Финансирование: Исследование выполнено при финансовой поддержке Министерства науки и высшего образования РФ по гранту № ФВНР-2022-0022 в Федеральном исследовательском центре ИЦиГ СО РАН и в рамках программы ICIG SB RAS Core Facility of Proteomic and metabolomic profiling of microorganisms (поддержана проектом № 075-15-2019-1662).

A GC-MS chemotaxonomic study on lipophilic compounds in the leaves of *Triticum aestivum* L.

Vasilieva A.R.[1, 2]*, Slynko N.M.[1, 2], Goncharov N.P.[1], Tatarova L.E.[1, 2], Kuibida L.V.[3], Peltek S.E.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Voevodsky Institute of Chemical Kinetics and Combustion, SB RAS, Novosibirsk, Russia*

* *vasilieva@bionet.nsc.ru*

Key words: *Triticum aestivum*; GC-MS; polymorphism; chemotaxonomy; metabolism

Motivation and Aim: Wheat *Triticum aestivum* L. is one of the major cereal crops on the globe. The number of wheat varieties and subspecies is enormous and continues to increase. Metabolic studies provide a deeper understanding of phenotypic diversity within the species polymorphism of *T. aestivum* and are key to chemotaxonomic classification and identification of species and subspecies based on their chemical composition. Determination of the chemical composition of lipophilic metabolites in green leaf extracts will provide an opportunity to establish relationships both within individual metabolic pathways and between these specific pathways. The realization of such a comparative study on a wide sample of *T. aestivum* cultivars and subspecies is the aim of this study.

Methods and Algorithms: A collection of 30 hexa- soft (*T. aestivum*) and its subspecies *T. aestivum* ssp. *petropavlovsky*i N.P. Gontsch., tetra- (*T. cartlicum* Nevski) No. 21, 24 and diploid (*T. monococcum* L.) No. 26, 27 wheat from the collection of the wheat

genetics sector of ICIG SB RAS was studied. For annotation of compounds, data were taken on the following systems: Pegasus 4D GCxGC-TOF MS (includes Agilent Technologies 7890B) and Agilent Technologies 6890 chromatograph with 5973 mass spectrophotometric detector. Hierarchical cluster analysis was used to visualize the differences between the studied leaf extracts of different wheat cultivars. For a more comprehensive coverage and in-depth analysis of the studied processes, we used multivariate analysis approaches, in particular, the principal component analysis (PCA) method.

Results: As a result of data interpretation on the basis of mass spectra analysis it was determined that aliphatic alcohols of linear structure and their derivatives have the highest content in the majority of extracts. Hydrocarbons, both linear and branched, were identified in the range C10–C40, among the linear ones C29, C31 predominate. A total of 16 saturated aliphatic acids, 5 of their esters, 5 unsaturated acids, 6 of their esters, and 2 lactones were identified. Aliphatic aldehydes and ketones are represented by 14 linear saturated and 5 unsaturated aldehydes, 4 linear and 2 branched ketones, 4-Hydroxy-2-butanone, and three β- and one γ-diketones. Hentriacontane-14,16-dione and its hydroxy derivative are reported in the literature to be a factor in wheat leaf ossification. Isoprene derivatives detected in the chromatograms of the samples include: monoterpenes (11) and cymols formed from them in the plant (2), diterpenes (17), sesquiterpenes (1), steroids and their derivatives (13), and their precursors (2). Identified compounds that include a polyprenoid fragment in their structure belong to the same group: vitamin E, tocopherols α, γ, their esters, tocospiro A and B, and 4,8,12,16-tetramethylheptadecan-4-olide. In addition, Geranylacetone, Farnesylacetone, and Geranyllinalool were identified in the chromatograms. Aromatic compounds identified in the extracts: phenols (13) and their derivatives (5), benzoic acid and its derivatives (3), aromatic systems conjugated to a carbonyl group (5) or to a double bond (3), condensed systems (5). In addition, Benzeneacetaldehyde was detected in the chromatograms. Peaks in mass spectra with Kovacs indices 2043, 2267, 2367, 2467, 2667, 2767, 2863 by the AMDIS, LECO ChromaTOF program are identified as either N-ethylamides or N,N-dimethylamides of tetradecanoic, hexadecanoic, heptadecanoic, octadecanoic, eicosanoic, genecosanoic and docosanoic acids respectively. Additional studies validate the structure of N-ethylamides of fatty acids.

No data are available on the metabolic synthesis of N-ethylamides of fatty acids. It can be assumed that this process involves the following steps: 1 – interaction of the fatty acid corresponding to the future amide with acetyl-CoA, proceeding with the consumption of ATP, 2 – reaction of the activated acid with alanine, with the formation of N-acylalanine fatty acid, 3 – reaction with the participation of decarboxylase enzyme and pyridoxal-5'-phosphate with the formation of N-ethylamide.

The PCA calculation using all the identified compounds allowed us to identify several clusters with possible chemotaxonomic similarities (Fig. 1). Samples with a predominant octacosanol content can be identified in a separate group located in the zone of negative PC1 values (claster 1), in contrast to samples with a predominant octacosanol and phytol acetates (claster 2). The samples with a predominance of fatty acid alkylamides (claster 3) are located separately.

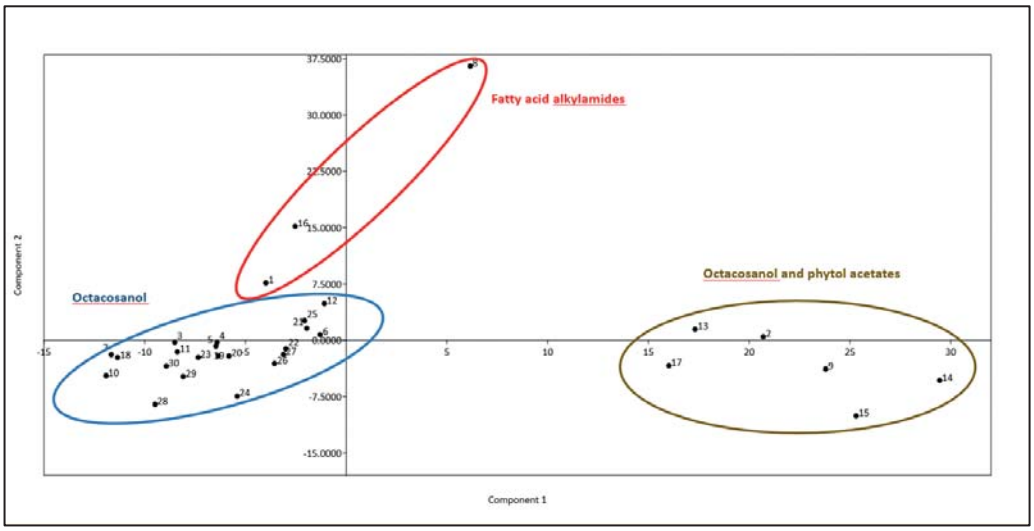


Fig. 1. PCA biplots based on all identified compounds in 30 wheat cultivars. The inner table contains the eigenvalues and percentage of variability for the first two principal components

Acylation of aliphatic alcohols, which are predominant in Cluster 1, into acetates (the primary components of Cluster 2) presumably occurs in response to changes in environmental humidity. This acylation blocks the hydrophilic fragments of the aliphatic alcohols, thereby imparting hydrophobic properties to the cuticle. As a result, this modification prevents excessive moisture from penetrating the leaf.

Conclusion: The study of compositions of lipophilic fractions of leaf extracts of 30 varieties of *Triticum aestivum* resulted in the presence of chemotaxonomic polymorphism.

Specific markers of these polymorphisms are: octacosanol, prevailing in the extracts of most of the studied varieties, as well as N-ethylamides of fatty acids, β-diketones and acetates of octacosanol and phytol.

N-ethylamides of fatty acids detected for the first time in large amounts in part of the sample extracts may play an important role in plant adaptation to abiotic stresses.

No correlations were found between cuticle composition and geographical origin.

Funding: The study was financially supported by the Ministry of Science and Higher Education of the Russian Federation under grant No. FVNR-2022-0022 at the Federal Research Center ICG SB RAS and under the ICG SB RAS Core Facility of Proteomic and metabolomic profiling of microorganisms (supported by project No. 075-15-2019-1662).

doi 10.18699/bgrs2024-3.2-05

Подходы искусственного интеллекта в задачах поиска молекулярных мишеней новых лекарственных препаратов

Васюченко Е.П.[1]*, Антонец Д.В.[1], Вяткин Ю.В.[1], Волчо К.П.[2]

[1] *Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем МГУ им. М.В. Ломоносова, Москва, Россия*

[2] *Отдел медицинской химии, Новосибирский институт органической химии им. Н.Н. Ворожцова Сибирского отделения Российской академии наук, Новосибирск, Россия*

** e.vasyuchenko@iai.msu.ru*

Ключевые слова: болезнь Паркинсона; разработка лекарств; молекулярный докинг

Мотивация и цель: исследовательская группа из Отдела медицинской химии Новосибирского института органической химии им. Н. Н. Ворожцова СО РАН обнаружила малые молекулы на основе монотерпеноидов, способные устранять нарушения на моделях болезни Паркинсона на мышах [1, 2]. Действие трех соединений не только снимало симптоматику болезни Паркинсона, но и приводило к выживанию и нормальному функционированию дофаминовых нейронов. Эти соединения включают диол (Проттремин), эпоксидиол, PA96. Помимо этого, для эпоксидиола известен стереоизомерный цис-1,2-эпоксид, а для PA96 существует его аналог, которые не имеет активности против болезни Паркинсона. Хотя Проттремин проходит клинические исследования, механизм антипаркинсонического действия этой группы соединений до сих пор не известен. Перед нами встала задача поиска белковой мишени для этих соединений.

Методы и алгоритмы: Для анализа было отобрано 69 структур белков, для которых показана ассоциация с болезнью Паркинсона. Структуры были взяты из открытой базы данных PDB Database (https://www.rcsb.org) в случае, если для белка не была получена кристаллическая структура, мы использовали структуру, предсказанную AlphaFold2 из базы данных UniProt (https://www.uniprot.org). Для поиска потенциальных мишеней был использован метод на основе диффузионной генеративной нейронной модели DiffDock [3]. Свободная энергия связывания лиганда с белком рассчитывалась методом MMPBSA.

Результаты: Предсказание сайтов связывания малых молекул на белке позволило выделить мишени, для которых был найден сайт связывания, и протестировать найденные сайты методом классического докинга. Для полученных комплексов рассчитывалась свободная энергия связывания лиганда с белком методом MMPBSA. Разница этой энергии между активными и неактивными соединениями служила критерием селективности. Те белки, для которых энергия связывания активных соединений была ниже, чем для неактивных были предложены в качестве потенциальных мишеней для экспериментальной проверки. Нами было выделено восемь потенциальных белковых мишеней, которые ждут экспериментальной проверки на способность связываться с активными молекулами.

Выводы: Разработанный нами подход к поиску молекулярных мишеней позволил выделить потенциальные таргетные белки, на которые нацелены молекулы антивопаркинсонического действия.

Финансирование: Данная работа является результатом исследовательского проекта, реализованного в рамках исследовательской программы Института искусственного интеллекта МГУ.

Artificial intelligence approaches to the problem of searching for molecular targets of new drugs

Vasyuchenko E.P.[1]*, Antonets D.V.[1], Vyatkin Yu.V.[1], Volcho K.P.[2]

[1] *Institute for Advanced Research on Artificial Intelligence and Intelligent Systems, Lomonosov Moscow State University, Moscow, Russia*

[2] *Department of Medicinal Chemistry, N. N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia*

** e.vasyuchenko@iai.msu.ru*

Key words: Parkinson's disease; drug development; molecular docking

Motivation and Aim: Research group from the Department of Medical Chemistry of the Novosibirsk Institute of Organic Chemistry named after. N.N. Vorozhtsova SB RAS discovered small molecules based on monoterpenoids that can eliminate disorders in mouse models of Parkinson's disease [1, 2]. The action of the three compounds not only relieved the symptoms of Parkinson's disease, but also led to the survival and normal functioning of dopamine neurons. These compounds include diol (Protremin), epoxydiol, PA96. In addition, a stereoisomeric cis-1,2-epoxide is known for epoxydiol, and its analogue exists for PA96, which has no activity against Parkinson's disease. Although Protremin is undergoing clinical trials, the mechanism of the antiparkinsonian action of this group of compounds is still unknown. We were faced with the task of finding a protein target for these compounds.

Methods and Algorithms: 69 protein structures that showed an association with Parkinson's disease were selected for analysis. Structures were taken from the public database PDB Database (https://www.rcsb.org) in case a protein crystal structure was not obtained, we used the structure predicted by AlphaFold2 from the UniProt database (https://www.uniprot.org). To search for potential targets, a method based on the diffusion generative neural model DiffDock [3] was used. The free energy of binding of the ligand to the protein was calculated by the MMPBSA method.

Results: Prediction of binding sites for small molecules on a protein made it possible to identify targets for which a binding site was found and to test the found sites using classical docking. For the resulting complexes, the free energy of binding of the ligand to the protein was calculated using the MMPBSA method. The difference in this energy between active and inactive compounds served as a criterion for selectivity. Those proteins for which the binding energy of active compounds was lower than for inactive compounds were proposed as potential targets for experimental testing. We have identified eight potential protein targets that await experimental testing for their ability to bind to active molecules.

Conclusion: The approach we developed to searching for molecular targets made it possible to identify potential target proteins that are targeted by antiparkinsonian molecules.

Funding: This work is an output of a research project implemented as part of the Research Program at the MSU Institute for Artificial Intelligence.

Список литературы/References

1. Kotliarova A. et al. A Newly Identified monoterpenoid-based small molecule able to support the survival of primary cultured dopamine neurons and alleviate MPTP-induced toxicity *in vivo*. *Molecules*. 27.23 (2022):8286. doi 10.3390/molecules27238286
2. Aleksandrova Yu. et al. Monoterpenoid epoxidiol ameliorates the pathological phenotypes of the rotenone-induced Parkinson's disease model by alleviating mitochondrial dysfunction. *International journal of molecular sciences*ю. 2023;24(6):5842. doi 10.3390/ijms24065842
3. Corso G. et al. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv*. 2022. doi 10.48550/arXiv.2210.01776

doi 10.18699/bgrs2024-3.2-07

Возможности оценки *in silico* фармакологического потенциала фитокомпонентов фармакопейных растений России

Ионов Н.*, Филимонов Д., Поройков В.
Институт биомедицинской химии им. В.Н. Ореховича, Москва, Россия
** ionov.nikita.serg@gmail.com*

Ключевые слова: Phyto4Health; PASS; природные соединения; *in silico*

Мотивация и цель: Природные соединения в сравнении с соединениями, доступными для синтеза, охватывают бóльшую область известного химического пространства и обладают более широким набором видов биологической активности [1]. Одним из источников таких соединений являются растения, используемые в традиционной медицине различных стран. Часто они характеризуются длительным опытом применения, что позволяет получить представление об их эффективности и безопасности. Ранее сведения о фитохимическом составе отдельных фармакопейных растений России были представлены в научных публикациях, монографиях и сборниках, что ограничивало возможности применения к ним компьютерных методов оценки биологической активности. В частности, отсутствовали возможности агрегации данных с учетом оценки достоверности предоставляемых результатов. Для решения этой проблемы нами разработана информационно-вычислительная платформа (ИВП) Phyto4Health. В рамках Phyto4Health агрегированы сведения о фитохимическом составе фармакопейных растений, а также представлены вычислительные инструменты для компьютерной оценки биологической активности и поиска структурных аналогов. Phyto4Health свободно доступна для широкого круга академических исследователей в сети Интернет (https://www.way2drug.com/phyto4health).
Целью нашей работы стала оценка применимости ИВП Phyto4Health для оценки фармакологического потенциала фитокомпонентов фармакопейных растений России с использованием сведений из свободно доступных больших баз данных структур природных соединений.
Методы и алгоритмы: Сведения о фитохимическом составе фармакопейных растений, а также молекулярных свойствах отдельных фитокомпонентов извлечены нами из текстов научных публикаций и свободно доступных баз данных. Структуры химических соединений были представлены в различных форматах, включая MOL и канонические SMILES.
Прогноз профилей биологической активности для фитокомпонентов осуществляется с использованием компьютерной программы PASS и специализированных баз знаний по зависимостям «структура-активность». В рамках ИВП Phyto4Health используются четыре базы знаний: фармакологические эффекты (PASS 2022 Refined [2], 358 видов), механизмы действия (PASS 2022 Refined, 1495 видов), нежелательные побочные эффекты (ADVERPred, 5 видов) [3], цитотоксическое действие в отношении нормальных и опухолевых клеточных линий (CLC-Pred 2.0, 438 видов) [4]. Для оценки показателей качества прогноза различных видов

биологической активности из базы данных LOTUS [5] было извлечено 191 886 структур природных соединений, которые потенциально могут быть выделены из растений. Для сформированного таким образом набора структур в базе данных PubChem был проведен поиск информации о результатах лабораторного тестирования их взаимодействия с молекулярными мишенями и клеточными линиями. Опубликованная в литературе информация сопоставлена с результатами компьютерного прогноза. На основе полученных данных были рассчитаны показатели точности, специфичности и чувствительности, характеризующие качество предсказаний.

Результаты: Phyto4Health предоставляет пользователю возможность получить сведения о фитохимическом составе 233 фармакопейных растений России. Для каждого из 3128 фитокомпонентов представлена информация о его свойствах, идентификаторах во внешних базах данных (PubChem, ChEMBL и ChEBI), а также результатах лабораторного тестирования взаимодействия с молекулярными мишенями. Ранее результаты компьютерного прогноза были подтверждены в рамках экспериментов, выполненных в сторонних организациях. В их числе оценка противоопухолевого эффекта фармацевтической композиции «Фитол-адаптоген» (НМИЦ онкологии им. Н.Н. Блохина) [6], оценка антиагрегантного и антикоагулянтного действия фитокомпонентов *Rubus chamaemorus* L. (СПХФУ) [7], а также оценка гепатопротекторного эффекта цикориевой и хлорогеновой кислот из *Cichorium intybus* L. (ФГБНУ ВИЛАР) [8]. Кроме того, с использованием ИВП Phyto4Health были проведены независимые эксперименты [9]. Однако отдельные примеры подтверждения результатов прогноза не позволяют получить комплексное представление об эффективности используемых нами методов для оценки биологической активности. Поэтому нами была проведена оценка эффективности с использованием данных из крупных репозиториев химических структур природных соединений. В ходе сопоставления экспериментальных данных и результатов компьютерного прогноза установлено, что средняя сбалансированная точность прогноза составила 0.7, чувствительность – 0.8, а специфичность – 0.6.

Выводы: На основе анализа данных из крупных химических репозиториев и результатов лабораторных экспериментов показана применимость используемых нами методов компьютерного прогноза биологической активности для выявления перспективных направлений исследования фитокомпонентов фармакопейных растений РФ. Разработанная нами платформа Phyto4Health предоставляет исследователям данные о фитокомпонентном составе фармакопейных растений России, а также вычислительные инструменты для оценки их фармакологического потенциала.

Финансирование: Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 гг.) (№ 122030100170-5).

In silico evaluation possibilities of Russian pharmacopoeial plants phytochemicals pharmacological potential

Ionov N.*, Filimonov D., Poroikov V.

Institute of Biomedical Chemistry, Moscow, Russia

** ionov.nikita.serg@gmail.com*

Key words: Phyto4Health; PASS; natural products; *in silico*

Motivation and Aim: Natural products cover a larger area of known chemical space and have a wide-ranging set of possible biological activities compared to available synthetic chemical compounds [1]. One of the sources of such chemical compounds is plants used in traditional medicine in various countries. They are often characterized by a long history of use, which provides insight into their efficacy and safety. However, information about the phytochemical composition of individual Pharmacopoeial plants of Russia has been presented in a number of scientific publications and compilations. Such situation restricts the possibilities of application of computer-based methods for analysis of biological activity. In particular, there was a lack of data aggregation capabilities to assess the validity of the provided results. To solve this problem, we developed the Phyto4Health Informational-Computational Platform. The platform aggregates data on the phytochemical composition of Pharmacopoeial plants and provides tools for computer-based evaluation of biological activity and search for structural analogues. Phyto4Health is freely available on the Internet (https://www.way2drug.com/phyto4health).

The aim of this study was to evaluate the applicability of this platform to assess the pharmacological potential of phytochemical of Pharmacopoeial plants in Russia, taking into account information from freely available large chemical databases of structures of natural products.

Methods and Algorithms: Information of the phytochemical composition of Pharmacopoeial plants and the molecular properties of individual phytochemicals were extracted from scientific publications and freely available databases. The structures of chemical compounds were presented in different formats, including MOL and Canonical SMILES.

The prediction of biological activity profiles for chemical structures of phytocomponents is performed using the PASS software [2] and the "structure-activity" relationships knowledge bases. The Phyto4Health uses four knowledge bases: pharmacological effects (PASS 2022 Refined [2], 358 species), mechanisms of action (PASS 2022 Refined, 1495 species), adverse effects (ADVERPred, 5 species) [3], cytotoxic effects against tumor and non-tumor cell lines (CLC-Pred 2.0, 438 species) [4]. To evaluate the quality prediction indicators of different types of biological activity, 191,886 structures of natural compounds that could potentially be isolated from plants were retrieved from the LOTUS database [5]. For the structures set, information on the results of laboratory testing of their interaction with molecular targets and cell lines was searched in the PubChem database. The information published in the literature was compared with the results of *in silico* prediction. On the basis of the obtained data, accuracy, specificity and sensitivity indices characterizing the quality of predictions were calculated.

Results: Phyto4Health provides the user with the opportunity to obtain information on the phytochemical composition of 233 Pharmacopoeial plants in Russia. For each of the 3,128 phytocomponents, information on the values of its molecular properties, identifiers in external databases (PubChem, ChEMBL and ChEBI), and the results of laboratory testing of interaction with molecular targets is presented. Formerly, the results of the computational prediction have been validated in laboratory experiments performed at third-party institutions. These included the evaluation of the antitumor effect of the pharmaceutical composition 'Phytoladaptogen' (N.N. Blokhin NMRC Oncology) [6], evaluation of the antiaggregant and anticoagulant effect of

phytocomponents of *Rubus chamaemorus* L. [7] (St. Petersburg State Chemical Pharmaceutical University), and evaluation of hepatoprotective effect of cichoric and chlorogenic acids from *Cichorium intybus* L. [8] (All-Russian Scientific Research Institute of Medicinal and Aromatic Plants). In addition, independent experiments were conducted based on information provided by Phyto4Health [9]. However, separate cases of prediction results validation do not provide a comprehensive view of the effectiveness of the methods we used to assess biological activity. Therefore, we evaluated the performance using data from large repositories of chemical structures of natural products. By comparing experimental data and computer prediction results, we found that the average balanced prediction accuracy was 0.7, sensitivity 0.8, and specificity 0.6.

Conclusion: Based on the analysis of data from large chemical repositories and the results of laboratory experiments, the applicability of our methods of computational prediction of biological activity for identifying promising ways to study phytocomponents of Pharmacopoeial plants of the Russian Federation is shown. The Phyto4Health platform provides Scientific Community with the reliable data on the phytochemical composition of Russian Pharmacopoeial plants, as well as computational tools to assess their pharmacological potential.

Funding: The study was performed in the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030) (No. 122030100170-5).

Список литературы/References

1. Chen Y. et al. Characterization of the chemical space of known and readily obtainable natural products. *J Chem Inf Model.* 2018;58(8):1518-1532
2. Poroikov V.V. Computer-Aided Drug Design: from Discovery of Novel Pharmaceutical Agents to Systems Pharmacology. *Biochem Moscow Suppl Ser B.* 2020;14(3):216-227
3. Ivanov S.M. et al. ADVERPred – web service for prediction of adverse effects of drugs. *J Chem Inf Model.* 2018;58(1):8-11
4. Lagunin A.A. et al. CLC-Pred 2.0: A Freely Available Web Application for In Silico Prediction of Human Cell Line Cytotoxicity and Molecular Mechanisms of Action for Druglike Compounds. *Int J Mol Sci.* 2023;24(2):1689
5. Rutz A. et al. The LOTUS initiative for open knowledge management in natural products research. *Elife*. 2022;11:e70780
6. Ionov N.S. et al. Possibilities of in Silico Estimations for the Development of the Pharmaceutical Composition Phytoladaptogene Cytotoxic for Bladder Cancer Cells. *Biochem Moscow Suppl Ser B.* 2021;15:290-300
7. Luzhanin V.G. et al. The Effect of Individual Compounds from Rubus chamaemoruson Hemostasis in vitro. *Drug Dev Regist.* 2024;13(1):149-158
8. Лупанова И.А. и др. Влияние цикориевой и хлорогеновой кислот из Cichorium intybus L. на активность цитохрома Р450 и глутатионтрансферазы. *Биофармацевтический журнал*. 2022;14(5):8-18
9. Filatov V. et al. Synergetic Effects of Aloe Vera Extract with Trimethylglycine for Targeted Aquaporin 3 Regulation and Long-Term Skin Hydration. *Molecules*. 2024;29(7):1540

doi 10.18699/bgrs2024-3.2-08

Поиск возможных молекулярных мишеней действия новых противотуберкулезных соединений

Калитванская М.А.[1], Макаров Д.А.[2], Хандажинская А.Л.[2], Анашкина А.А.[2, 3]*

[1] *Волгоградский медицинский университет, Волгоград, Россия*

[2] *Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

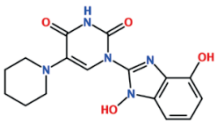
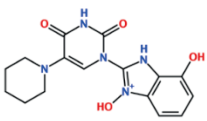
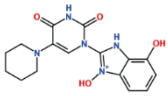
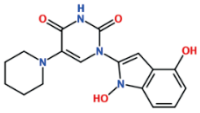
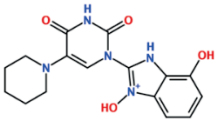
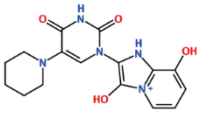
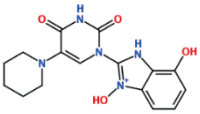
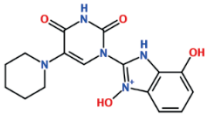
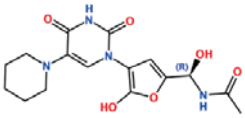
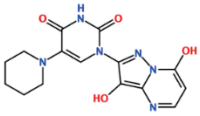
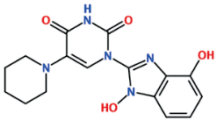
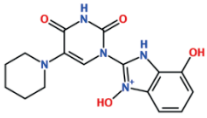
[3] *Сеченовский университет, Москва, Россия*

* *anastasia.a.anashkina@mail.ru*

Ключевые слова: мишени; туберкулез; противотуберкулезные соединения

Мотивация и цель: На сегодняшний день туберкулез является одной из главных проблем в мире, так как умерло и продолжает умирать людей от туберкулеза больше, чем от СПИДа, малярии и других болезней вместе взятых. По данным Всемирной организации здравоохранения (ВОЗ), на 2020 г. туберкулезом в мире было инфицировано около 1 млрд человек, из которых 70 млн умерло [1]. Несмотря на то что туберкулез излечим, главная угроза состоит в том, что микобактерии быстро приобретают устойчивость к лекарственным препаратам – со скоростью 0.5 SNPs на геном в год [2, 3]. Устойчивость микобактерий к римфапицину (РИФ) в 2022 г. была выявлена у более чем 450 тыс. пациентов, а у 1.3 млн человек развился туберкулез, устойчивый к изониазиду (ИНГ). Механизмы устойчивости к этим препаратам уже хорошо известны, на данный момент составлен каталог мутаций и их связи с фенотипической лекарственной устойчивостью, что облегчает разработку лекарственных средств [4]. Однако проблема до конца не решена, и устойчивость возникает ко всем появляющимся препаратам. Целью нашей работы является проведение скрининга структурно близких соединений уже изученных ранее соединений с подтвержденной активностью и *in silico* изучение их взаимодействия с рядом белков *M. tuberculosis*.
Методы и алгоритмы: В качестве основы для разработки новых противотуберкулезных соединений мы взяли ряд соединений с известной противотуберкулезной активностью и методом Scaffold Replacement в программе MOE создали библиотеки химических соединений с последующим докингом к семи белкам *M. tuberculosis*: 6-hydroxymethyl-7,8-dihydropteroate synthase (1eye [5]), isocitrate lyase (1f8i, 1f61 [6]), Alkylperoxidase (1gu9, 1lw1 [7, 8]), Type II Dehydroquinase (1h0r [9]), 3-dehydroquinate dehydratase (1h0s, 1h05 [10, 11]), diaminopimelate dicarboxylase (1hkv [12]), Acyl Carrier Protein (1klp [13]).
Результаты: Найдено, что соединение 1-(1,4-дигидрокси-1Н-бензо[d]имидазол-2-ил)-5-(пиперидин-1-ил)пиримидин-2,4(1Н,3Н)-дион, возможно, взаимодействует с белком 6-hydroxymethyl-7,8-dihydropteroate synthase (1eye) (табл. 1). Кроме того, в протонированной по атому азота в положении 1 имидазольного фрагмента форме данное соединение может взаимодействовать с белком isocitrate lyase (1f8I) и 3-dehydroquinate dehydratase (1h0s). Возможный сайт связывания 6-hydroxymethyl-7,8-dihydropteroate synthase (1eye) включает остатки: Arg 253, Gly 50, Gly 22, Val 107, His 255, Asp 86, Asp 21, Phe 182, Ser 20, Gly 209, Leu 207, Lys 213, Asn 105, Asp 177, Met 130, Asn 13, Val 11, Ser 211, Val 128, Leu 180, Thr 15, Asp 17 и также обладает лучшим результатом по афинности связывания (–8.0 ккал/моль) (рис. 1).

Таблица 1. Соединения, показавшие наилучшую афинность в эксперименте по докингу к белкам *M. tuberculosis*

1eye	O=C1N(c2n(O)c3c(c(O)ccc3)n2)C=C(N2CCCCC2)C(=O)N1 	-8.0	1h0r	O=C(N[C@H](O)c1oc(O)c(N2C(=O)NC(=O)C(N3CCCCC3)=C2)c1)C 	-6.8
1f8i	O=C1N(c2[n+](O)c3c(c(O)ccc3)[nH]2)C=C(N2CCCCC2)C(=O)N1 	-7.6	1hkv	O=C1N(c2n(O)c3c(c(O)ccc3)c2)C=C(N2CCCCC2)C(=O)N1 	-6.6
1h0s	O=C1N(c2[n+](O)c3c(c(O)ccc3)[nH]2)C=C(N2CCCCC2)C(=O)N1 	-7.0	1klp	O=C1N(c2n(O)c3c(c(O)ccc3)c2)C=C(N2CCCCC2)C(=O)N1 	-6.6
1lwl	O=C1N(c2n(O)c3c(c(O)ccc3)[n+H]2)C=C(N2CCCCC2)C(=O)N1 	-6.9	1f61	O=C1N(c2[n+](O)c3c(c(O)ccc3)[nH]2)C=C(N2CCCCC2)C(=O)N1 	-6.4
1gu9	O=C(N[C@H](O)c1oc(O)c(N2C(=O)NC(=O)C(N3CCCCC3)=C2)c1)C 	-6.8	1h05	O=C1N(c2n(O)c3[n+](c(O)ccc3)c2)C=C(N2CCCCC2)C(=O)N1 	-6.4
1eye	O=C1N(c2n(O)c3c(c(O)ccc3)n2)C=C(N2CCCCC2)C(=O)N1 	-8.0	1h0r	O=C(N[C@H](O)c1oc(O)c(N2C(=O)NC(=O)C(N3CCCCC3)=C2)c1)C 	-6.8

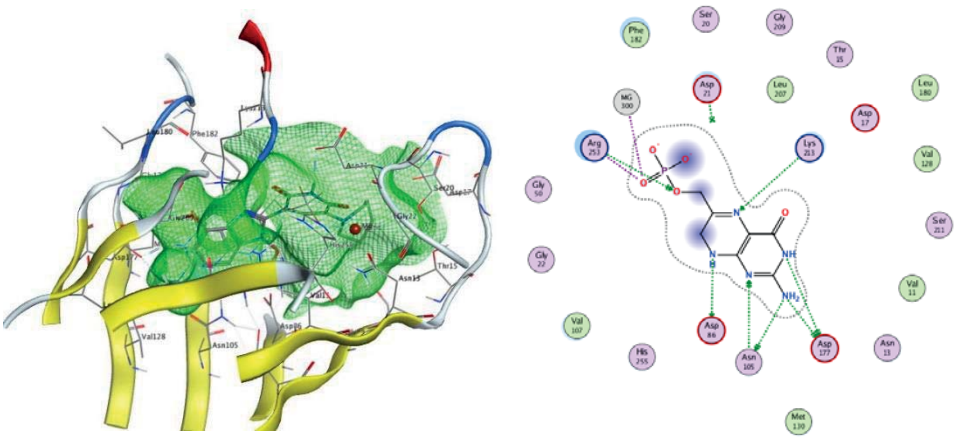


Рис. 6. Потенциальный сайт связывания найденного соединения в белке 6-hydroxymethyl-7,8-dihydropteroate synthase (1eye) включает остатки Arg 253, Gly 50, Gly 22, Val 107, His 255, Asp 86, Asp 21, Phe 182, Ser 20, Gly 209, Leu 207, Lys 213, Asn 105, Asp 177, Met 130, Asn 13, Val 11, Ser 211, Val 128, Leu 180, Thr 15, Asp 17

Выводы: Методами молекулярного моделирования найдено три потенциально активных химических соединения с противотуберкулезной активностью.
Финансирование: Исследование поддержано грантом Российского научного фонда № 23-14-00106 «Новое поколение биоцидов с широким спектром биологического действия».

Search for possible molecular targets of new anti-tuberculosis compounds

Kalitvanskaya M.A.[1], Makarov D.A.[2], Khandazhinskaya A.L.[2], Anashkina A.A.[2, 3]*

[1] *Volgograd Medical University, Volgograd, Russia*

[2] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

[3] *Sechenov University, Moscow, Russia*

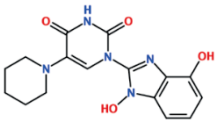
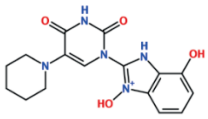
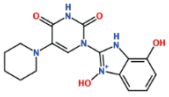
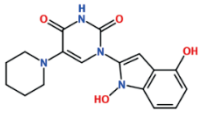
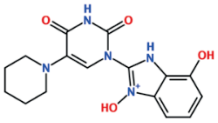
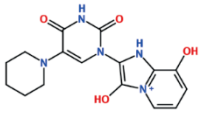
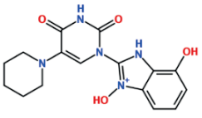
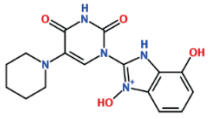
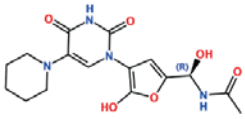
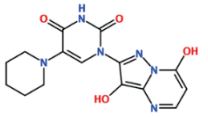
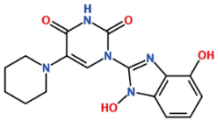
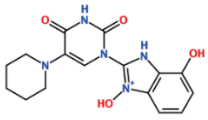
* *anastasia.a.anashkina@mail.ru*

Key words: targets; tuberculosis; anti-tuberculosis compounds

Motivation and Aim: Today, tuberculosis is one of the main problems in the world, since more people have died and continue to die from tuberculosis than from AIDS, malaria and other diseases combined. According to the World Health Organization (WHO), in 2020, about 1 billion people were infected with tuberculosis in the world, 70 million of whom died [1]. Although tuberculosis is curable and the main threat is that mycobacteria are rapidly acquiring drug resistance at a rate of 0.5 SNPs per genome per year [2, 3]. Mycobacterial resistance to rimphapicin (RIF) was detected in more than 450 thousand patients in 2022, and 1.3 million people developed isoniazid-resistant tuberculosis (INR). The mechanisms of resistance to these drugs are already well known; a catalog of mutations and their relationship with phenotypic drug resistance has now been compiled,

which facilitates the development of drugs [4]. However, the problem has not been completely solved, and resistance occurs to all emerging drugs. The purpose of this work is to screen structurally similar compounds of previously studied compounds with confirmed activity and in silico study of their interaction with a number of *M. tuberculosis* proteins.

Table 1. Compounds that showed the best affinity in the docking experiment for *M. tuberculosis* proteins

1eye	O=C1N(c2n(O)c3c(c(O)ccc3)n2)C=C(N2CCCCC2)C(=O)N1 	-8.0	1h0r	O=C(N[C@H](O)c1oc(O)c(N2C(=O)NC(=O)C(N3CCCCC3)=C2)c1)C 	-6.8
1f8i	O=C1N(c2[n+](O)c3c(c(O)ccc3)[nH]2)C=C(N2CCCCC2)C(=O)N1 	-7.6	1hkv	O=C1N(c2n(O)c3c(c(O)ccc3)c2)C=C(N2CCCCC2)C(=O)N1 	-6.6
1h0s	O=C1N(c2[n+](O)c3c(c(O)ccc3)[nH]2)C=C(N2CCCCC2)C(=O)N1 	-7.0	1klp	O=C1N(c2n(O)c3c(c(O)ccc3)c2)C=C(N2CCCCC2)C(=O)N1 	-6.6
1lwl	O=C1N(c2n(O)c3c(c(O)ccc3)[n+H]2)C=C(N2CCCCC2)C(=O)N1 	-6.9	1f61	O=C1N(c2[n+](O)c3c(c(O)ccc3)[nH]2)C=C(N2CCCCC2)C(=O)N1 	-6.4
1gu9	O=C(N[C@H](O)c1oc(O)c(N2C(=O)NC(=O)C(N3CCCCC3)=C2)c1)C 	-6.8	1h05	O=C1N(c2n(O)c3[n+](c(O)ccc3)c2)C=C(N2CCCCC2)C(=O)N1 	-6.4
1eye	O=C1N(c2n(O)c3c(c(O)ccc3)n2)C=C(N2CCCCC2)C(=O)N1 	-8.0	1h0r	O=C(N[C@H](O)c1oc(O)c(N2C(=O)NC(=O)C(N3CCCCC3)=C2)c1)C 	-6.8

Methods and Algorithms: As a basis for the development of new anti-tuberculosis compounds, we took a number of compounds with known anti-tuberculosis activity and created libraries of chemical compounds using the Scaffold Replacement method in the MOE program, followed by docking to seven *M. tuberculosis* proteins (6-hydroxymethyl-7,8-dihydropteroate synthase (1eye [5]), isocitrate lyase (1f8i, 1f61 [6]), Alkylperoxidase(1gu9, 1lw1 [7], [8]), Type II Dehydroquinase (1h0r [9]), 3-dehydroquinate dehydratase (1h0s, 1h05 [10], [11]), diaminopimelate dicarboxylase (1hkv [12]), Acyl Carrier Protein (1klp [13]).

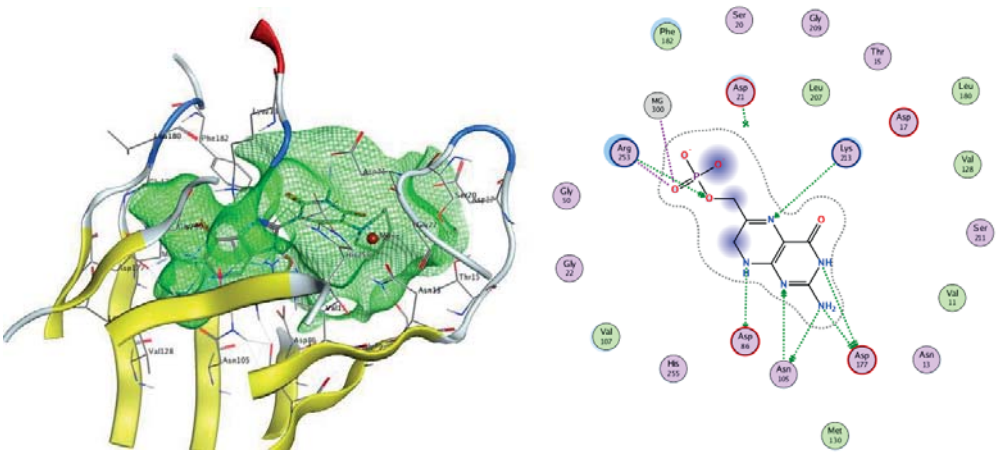


Fig. 1. Potential binding site of the found compound in the protein 6-hydroxymethyl-7,8-dihydropteroate synthase (1eye) includes residues Arg 253, Gly 50, Gly 22, Val 107, His 255, Asp 86, Asp 21, Phe 182, Ser 20, Gly 209, Leu 207, Lys 213, Asn 105, Asp 177, Met 130, Asn 13, Val 11, Ser 211, Val 128, Leu 180, Thr 15, Asp 17

Results: It was found that the compound 1-(1,4-dihydroxy-1H-benzo[d]imidazol-2-yl)-5-(piperidin-1-yl)pyrimidin-2,4(1H,3H)-dione may interacts with the 6-hydroxymethyl-7,8-dihydropteroate synthase (1eye) protein (Table 1). In addition, in the form protonated at the nitrogen atom in position 1 of the imidazole fragment, this compound can interact with the protein isocitrate lyase (1f8I) and 3-dehydroquinate dehydratase (1h0s). Possible binding site for 6-hydroxymethyl-7,8-dihydropteroate synthase (1eye) includes residues: Arg 253, Gly 50, Gly 22, Val 107, His 255, Asp 86, Asp 21, Phe 182, Ser 20, Gly 209, Leu 207, Lys 213, Asn 105, Asp 177, Met 130, Asn 13, Val 11, Ser 211, Val 128, Leu 180, Thr 15, Asp 17 and also has the best binding affinity (–8.0 kcal/mol) (Fig. 1).

Conclusions: Using molecular modeling methods, three potentially active chemical compounds with anti-tuberculosis activity were found.

Funding: The study was supported by the Russian Science Foundation grant No. 23-14-00106 “New generation of biocides with a wide spectrum of biological action”.

Список литературы/References

1. Кривохиж В. Туберкулез. Современный взгляд на лечение и профилактику. Litres, 2019
2. CDCTB. Tuberculosis (TB) – Drug-Resistant TB [Electronic resource]. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/tb/topic/drtb/default.htm (accessed: 14.05.2024)
3. Comín J. et al. Estimation of the mutation rate of Mycobacterium tuberculosis in cases with recurrent tuberculosis using whole genome sequencing. *Sci Rep*. 2022;12(1):16728
4. Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance, 2nd ed [Electronic resource]. URL: https://www.who.int/publications-detail-redirect/9789240082410 (accessed: 14.05.2024)
5. Baca A.M. et al. Crystal structure of *Mycobacterium tuberculosis* 6-hydroxymethyl-7,8-dihydropteroate synthase in complex with pterin monophosphate: new insight into the enzymatic mechanism and sulfa-drug action. *J Mol Biol*. 2000;302(5):1193-1212
6. Sharma V. et al. Structure of isocitrate lyase, a persistence factor of Mycobacterium tuberculosis. *Nat Struct Mol Biol*. 2000;7(8):663-668
7. Nunn C.M. et al. The Crystal Structure of Mycobacterium tuberculosisAlkylhydroperoxidase AhpD, a Potential Target for Antitubercular Drug Design. *J Biol Chem*. 2002;277(22):20033-20040
8. Koshkin A. et al. The Mechanism of Mycobacterium tuberculosis Alkylhydroperoxidase AhpD as Defined by Mutagenesis, Crystallography, and Kinetics. *J Biol Chem*. 2003;278(32):29502-29508
9. Lence E. et al. Mechanistic Basis of the Inhibition of Type II Dehydroquinase by (2S)- and (2R)-2-Benzyl-3-dehydroquinic Acids. *ACS Chem Biol*. 2013;8(3):568-577
10. Peón A., Otero J.M., Tizón L. Understanding the Key Factors that Control the Inhibition of Type II Dehydroquinase by (*2R*)-2-Benzyl-3-dehydroquinic Acids. *ChemMedChem*. 2010;5(10):1726-1733
11. Evans L.D.B., Roszak A.W., Noble L.J. et al. Specificity of substrate recognition by type II dehydroquinases as revealed by binding of polyanions. *FEBS Lett*. 2002;530(1-3):24-30. doi 10.1016/S0014-5793(02)03346-X
12. Gokulan K. et al. Crystal Structure of Mycobacterium tuberculosis Diaminopimelate Decarboxylase, an Essential Enzyme in Bacterial Lysine Biosynthesis. *J Biol Chem*. 2003;278(20):18588-18596
13. Wong H.C. et al. The Solution Structure of Acyl Carrier Protein from Mycobacterium tuberculosis. *J Biol Chem*. 2002;277(18):15874-15880

doi 10.18699/bgrs2024-3.2-09

Прогноз противовирусной активности низкомолекулярных соединений на основе протеохемометрики

Карасев Д.А.*, Соболев Б.Н., Филимонов Д.А., Лагунин А.А., Тарасова О.А., Поройков В.

Институт биомедицинской химии им. В.Н. Ореховича, Москва, Россия

** d.karasev@ibmc.msk.ru*

Ключевые слова: протеохемометрика; низкомолекулярные лекарственноподобные соединения; хемоинформатика; анализ взаимосвязи «структура-активность»; (Q)SAR; противовирусные препараты; вирусные инфекции

Мотивация и цель: Оценка взаимосвязи «структура–активность» ((Q)SAR) на основе представления молекул в виде химических структур широко используется при компьютерном поиске новых лекарственных соединений. Методы протеохемометрики (PCM) позволяют расширить область применимости моделей взаимосвязи «структура-активность», поскольку при построении таких моделей также используются данные о структуре белков-мишеней. Таким образом, становится возможным осуществлять прогноз для белков с неизвестным спектром лигандов. Поскольку белки-мишени, кодируемые генами различных вирусов, могут проявлять значимое структурное сходство [1], использование PCM представляется весьма перспективным при поиске действенных лекарств при новых вирусных угрозах. Цель настоящей работы – разработка подхода PCM и его тестирование при прогнозе активности низкомолекулярных соединений в отношении протеаз коронавирусов, энтеровирусов, флавивирусов и вирусов гепатита С.

Методы и алгоритмы: Мы разработали оригинальный метод PCM, основанный на комбинированном описании пар белок-лиганд. Метод был протестирован на данных, представляющих вирусные протеазы из надсемейства химотрипсина и их ингибиторы. Данные о соответствующих взаимодействиях были извлечены с помощью оригинальной процедуры из базы данных ChEMBL. Множественное выравнивание белков было построено на основе попарных трехмерных выравниваний. Это позволило рассчитать дескрипторы белков с учетом областей связывания лигандов [2]. Для описания лигандов были применены дескрипторы атомных окрестностей (MNA-дескрипторы) [3].

Результаты: Тестирование набора лигандов в режиме (Q)SAR показало точность прогноза около 0.95 при валидации методом исключения по одному. Модели PCM, построенные с использованием комбинированных дескрипторов, были тестированы с использованием процедуры, в которой реализовано исключение пар белок-лиганд. Таким образом, было смоделировано предсказание взаимодействия новых лигандов и новых мишеней. Полученные значения точности при этом достигли 0.89.

Выводы: Таким образом, подход PCM показывает некоторое снижение точности по сравнению с (Q)SAR, но обеспечивает эффективное предсказание при отсутствии надежных данных о взаимодействии низкомолекулярных соединений с белками-мишенями соответствующих вирусов.

Финансирование: Исследование выполнено при поддержке программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 гг.) (№ 124050800018-9).

Prediction of antiviral activity of druglike compounds based on proteochemometrics

Karasev D.A.*, Sobolev B.N., Filimonov D.A., Lagunin A.A., Tarasova O.A., Poroikov V.V.

Institute of Biomedical Chemistry, Moscow, Russia

** w.dmitrykarasev@gmail.com*

Key words: proteochemometrics; small molecule drug-like compounds; chemoinformatics; structure-activity relationship analysis; (Q)SAR; antiviral drugs; viral infections

Motivation and Aim: An analysis of structure-activity relationships ((Q)SAR) based on the chemical structures of molecules is widely used in the computer search for new drug compounds. Proteochemometrics (PCM) methods expand the applicability area of structure–activity relationship models, as such models can use data on the structures of both protein targets and ligands, which is beneficial for predicting activity spectra for proteins with an unknown spectrum of ligands. Since target proteins encoded by genes of different viruses can exhibit significant structural similarity [1], the use of PCM appears is a promising in the search for effective drugs against new viral threats. The purpose of this work is to develop the PCM approach and test it for predicting the activity of small molecular compounds against proteases of coronaviruses, enteroviruses, flaviviruses and hepatitis C viruses.
Methods and Algorithms: We developed an original PCM method based on the combined description of protein-ligand pairs. The method was tested on data representing viral proteases from the chymotrypsin superfamily and their inhibitors. Data on relevant interactions were extracted using an original procedure from the ChEMBL database. The multiple protein alignment was built based on the pairwise 3D alignments. It enabled the calculation of protein descriptors, taking into account the regions of ligand binding [2]. Atomic neighborhood descriptors (MNA descriptors) were used to describe ligands [3].
Results: Testing a set of ligands in (Q)SAR mode with leave-one-out cross validation was characterized by a prediction accuracy of about 0.95. PCM models built using the combination descriptors were tested using a procedure involving the exclusion of protein-ligand pairs. Thus, the prediction of interactions between new ligands and new targets was modeled. The accuracy values obtained in PCM mode reached 0.89.
Conclusion: Thus, the PCM approach shows insignificant decrease in accuracy compared to (Q)SAR, but provides effective prediction in the absence of reliable data on the interaction of small molecules with the target proteins of the corresponding viruses.
Funding: The study was performed in the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030) (No. 124050800018-9).

Список литературы/References

1. Monttinen H.A.M., Ravantti J.J., Poranen M.M. Structural comparison strengthens the higher-order classification of proteases related to chymotrypsin. *PLoS One*. 2019;14(5):e0216659
2. Karasev D.A., Sobolev B.N., Filimonov D.A., Lagunin A.A. Prediction of viral protease inhibitors using proteochemometrics approach. *Comput Biol Chem*. 2024;110:108061
3. Lagunin A., Stepanchikova A., Filimonov D., Poroikov V. PASS: prediction of activity spectra for biologically active substances. *Bioinformatics*. 2000;16(8):747-748

doi 10.18699/bgrs2024-3.2-10

Идентификация РНК-связывающего участка белка с доменом холодового шока CspA из *Mycobacterium tuberculosis* методами молекулярного докинга и молекулярной динамики

Панкратова П.*, Леконцева Н., Смольянова Н., Никулин А.
Институт белка РАН, Пущино, Россия
** pankratova.p.y@gmail.com*

Ключевые слова: белки холодового шока; CspA; малые регуляторные РНК; РНК-белковые комплексы; *Mycobacterium tuberculosis*

Мотивация и цель: Транс-кодируемые малые регуляторные РНК (мрРНК) являются посттранскрипционными регуляторами трансляции, что позволяет бактериями *быстро адаптироваться к изменениям окружающей среды и воздействиям различных стрессов* [1, 2].

Бактерия *Mycobacterium tuberculosis* является возбудителем туберкулеза человека и входит в десятку наиболее важных причин смертности во всем мире. С помощью биоинформатического анализа в геноме *M. tuberculosis* выявлено около 2000 участков ДНК, которые могут кодировать малые нетранслируемые РНК [3]. Однако возможное участие нетранслируемых РНК в регуляции экспрессии генов и развитии стадий инфекции *M. tuberculosis* исследованы слабо, и функционально охарактеризованы только 8 мрРНК.

Взаимодействию между мрРНК и их мишенями часто способствуют белковые РНК-шапероны. В *M. tuberculosis* не обнаружены классические бактериальные РНК-шапероны Hfq и ProQ. Претендентом на роль РНК-шаперонов в *M. tuberculosis* могут служить белки с доменом холодового шока Csp (Cold shock protein), которые выполняют широкий спектр биологических функций в клетках различных организмов [4–6].

В нашей работе была поставлена цель создать модель взаимодействия белка CspA из *M. tuberculosis* (далее *Mtb*CspA) с мрРНК методами молекулярного докинга и молекулярной динамики.

Методы и алгоритмы: Для предсказания структуры белка *Mtb*CspA использовали аминокислотную последовательность из базы UNIPROT. Поиск гомологов проводился при помощи метода BLAST. Пространственная структура белка *Mtb*CspA промоделирована с помощью программы AlphaFolf2. Полученная модель *Mtb*CspA подвергалась стабилизации с помощью молекулярной динамики в GROMACS с силовым полем CHARMM в сочетании с явной моделью воды TIP3P в 0.1 M растворе NaCl при температуре 300 K. Длина траектории составила 300 нс.

Для формирования моделей комплексов *Mtb*CspA с РНК были взяты как короткие последовательности одноцепочечной РНК, так и несколько шпилечных структур мрРНК. Вторичная структура исследуемых мрРНК была рассчитана программами RNAfold и MXfold2. Пространственная структура шпилечных РНК была

промоделирована в программе VfoldLA. Структуры РНК-белковых комплексов были получены макромолекулярным докингом в программе HDOCK.
Результаты: Проанализирован ряд комплексов между *Mtb*CspA и фрагментами РНК длиной от 12 до 20 рибонуклеотидов. Нами идентифицированы две области связывания исследуемых РНК. Первый сайт представляет собой область аминокислотных остатков Trp8, Phe28, Arg57-Gly58-Pro59 и соответствует общему мотиву для семейства белков холодового шока, ранее описанному для белков группы CSP из *Bacillus subtilis* и *Escherichia coli*. Второй участок, в области петель белка β3-β4 и β4-β5, ранее не был описан и идентифицирован нами впервые. Одноцепочечные РНК связывались с поверхностью белка *Mtb*CspA как с первым участком, так и со вторым, причем были стабильны на протяжении траектории MD 300 нс. Шпилечные РНК имеют преимущественное сродство ко второй области, однако, такие комплексы менее стабильны по сравнению с комплексами оцРНК-*Mtb*CspA.
Выводы: На поверхности белка CspA из *M. tuberculosis* идентифицированы две предполагаемые области связывания одноцепочечных РНК и шпилечных РНК, причем вторая область взаимодействия нами была идентифицирована впервые.
Финансирование: Исследование выполнено при финансовой поддержке РНФ в рамках научного проекта № 24-24-00071.

Identification of the RNA-binding region of the cold shock protein CspA from *Mycobacterium tuberculosis* using molecular docking and molecular dynamics methods

Pankratova P.*, Lekontseva N., Smolyanova N., Nikulin A.
Institute of Protein Research, RAS, Pushchino, Russia
** pankratova.p.y@gmail.com*

Key words: cold shock proteins; CspA, sRNA; RNA-protein complexes; *Mycobacterium tuberculosis*

Motivation and Aim: Transcoded small regulatory RNAs (sRNAs) are post-transcriptional translational regulators that allow bacteria to adapt rapidly to environmental changes and various stresses [1, 2]. *Mycobacterium tuberculosis* is the causative agent of human tuberculosis and ranks among the top ten leading causes of mortality worldwide. Through bioinformatics analysis, approximately 2000 DNA loci have been identified in the *M. tuberculosis* genome that have the potential to encode small non-coding RNAs [3]. However, the potential involvement of non-coding RNAs in regulating gene expression and the developmental *M. tuberculosis* infection has been poorly investigated, and only 8 sRNAs have been functionally characterized.
The interaction between sRNAs and their targets is often facilitated by proteins called RNA chaperones. Notably, the classical RNA chaperones Hfq and ProQ have not been identified in *M. tuberculosis*. Potential candidates for RNA chaperone in *M. tuberculosis* could be proteins containing the Cold Shock Protein (Csp) domain, which perform a wide range of biological functions in cells of various organisms [4–6]. Our work aimed to create a model of the interaction of the protein CspA from *M. tuberculosis* (hereinafter *Mtb*CspA) with sRNA using molecular docking and molecular dynamics methods.

Methods and Algorithms: To predict the structure of the *Mtb*CspA protein, the amino acid sequence from the UNIPROT database was used. The search for homologs was carried out using the BLAST method. The 3D structure of the *Mtb*CspA protein was modeled using the AlphaFolf2 program. The resulting *Mtb*CspA model was subjected to molecular dynamics (MD) stabilization in GROMACS with the CHARMM force field coupled with an explicit TIP3P water model in 0.1 M NaCl solution at 300 K. The trajectory length was 300 ns.

To model the *Mtb*CspA complexes with RNA, both short sequences of single-stranded RNA and several hairpin structures of sRNA were taken. The secondary structure of the studied sRNAs was calculated using the RNAfold and MXfold2 programs. The spatial structure of hairpin RNAs was modeled using the VfoldLA program. The structures of RNA-protein complexes were obtained by macromolecular docking using the HDOCK program.

Results: A number of complexes between *Mtb*CspA and RNA fragments from 12 to 20 ribonucleotides in length were analyzed. We have identified two RNA-binding sites on the protein. The first site is a region of amino acid residues Trp8, Phe28, Arg57-Gly58-Pro59 and corresponds to a common motif for the cold shock protein family previously described for CSP proteins from *Bacillus subtilis* and *Escherichia coli*. The second site, including protein loops β3-β4 and β4-β5, has not been previously described and was identified by us for the first time. Single-stranded RNAs bound to the surface of the *Mtb*CspA protein both in the first and the second RNA-binding sites, and were stable throughout the MD trajectory of 300 ns. Hairpin RNAs have a preferential affinity for the second RNA-binding site; however, such complexes are less stable compared to ssRNA-*Mtb*CspA complexes.

Conclusions: We identified two putative RNA-binding sites on the surface of the CspA protein from *M.* tuberculosis, with the second RNA binding site being described for the first time.

Both *Mtb*CspA RNA-binding sites had an affinity for both single-stranded RNA and hairpin RNA.

Funding: The study is supported by the RSF (24-24-00071).

Список литературы/References

1. Gong C.C., Klumpp S. Modeling sRNA-Regulated Plasmid Maintenance. *PLoS One.* 2017;12(1):e0169703. doi 10.1371/journal.pone.0169703
2. Jørgensen M.G., Pettersen J.S., Kallipolitis B.H. sRNA-mediated control in bacteria: An increasing diversity of regulatory mechanisms. *Biochim Biophys Acta Gene Regul Mech.* 2020;1863(5):194504. doi 10.1016/j.bbagrm.2020.194504
3. Острик А.А., Ажикина Т.Л, Салина Е.Г. Малые некодирующие РНК и их роль в патогенезе Mycobacterium tuberculosis. *Успехи биологической химии.* 2021;61:229-252
[Ostrik A.A., Azhikina T.L., Salina E.G. Small non-coding RNAs and their role in the pathogenesis of Mycobacterium tuberculosis. *Advances in Biological Chemistry.* 2021;61:229-252 (in Russian)]
4. Corley M., Burns M.C., Yeo G.W. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Mol Cell.* 2020;78(1):9-29. doi 10.1016/j.molcel.2020.03.011
5. Roy A., Ray S. An in-silico study to understand the effect of lineage diversity on cold shock response: unveiling protein-RNA interactions among paralogous CSPs of E. coli. *3 Biotech.* 2023;13(7):236. doi 10.1007/s13205-023-03656-2
6. Zhang P., Wu W., Chen Q., Chen M. Non-Coding RNAs and their Integrated Networks. *J Integr Bioinform*. 2019;16(3):20190027. doi 10.1515/jib-2019-0027

doi 10.18699/bgrs2024-3.2-11

Химическая модификация противомикробных комплексов органическими молекулами

Плотникова Ю.*, Барышева Е., Бибарцева Е., Баранова О.
[1] *ФГБОУ ВО Оренбургский государственный университет, Оренбург, Россия*
* *shik8mail@mail.ru*

Ключевые слова: галловая кислота, молекулярно-электростатический потенциал, квантово-химическое моделирование, синергетическое взаимодействие

Цель: Целью проведенного исследования было определение синергического антибактериального эффекта совместного применения антибиотиков и вторичных метаболитов фенольного происхождения, а также определение характера их взаимодействия.
Материалы и методы: Материалом исследования стали клинические изоляты P. aeruginosa и S. aureus. Исследовались, предварительно экстрагированные фенольные компоненты 10 лекарственных растений: Herba Achilleae millefolii; Folia Eucalypti viminalis; Urticae folia; Cortex Quercus; Herba; Camellia; Ribes nigrum L.; Sorbus aucuparia L.; Allium sativum; Dianthus в комплексе с антибиотическими препаратами цефтазидимом и фосфомицином.
Апробация комбинированного действия антибиотиков и фенолсодержащих экстрактов проводилась методом агаровых лунок.
При проведении квантово-химического расчета сначала была проведена оптимизация молекул исследуемых веществ в небольшом приближении PBE/3-21+G. Осуществлялся конфигурационный поиск оптимального пространственного расположения функциональных групп, и на следующем этапе – выполнялся расчет исследуемых молекул, с построением карт молекулярно-электростатического потенциала (МЭП). Рассчитанные на первом этапе молекулы, с учетом МЭП, попарно соединялись и оптимизировались на основе теории функционала плотности в приближении PBE/3-21+G.
Спектры чистых веществ и предполагаемых межмолекулярных комплексов получали с использованием Фурье – спектрометра «ИНФРАЛЮМ ФТ-02» и ИК-Фурье спектрометра Bruker Alpha (без приставки НПВО).
Результаты: Изначально, были проведены исследования противомикробной активности, в условиях in vitro, при совместном введении в среду вторичных метаболитов фенольного происхождения и антибиотиков (фосфомицина и цефтазидима). В отношении S. aureus наиболее выраженный синергический эффект отмечен для Folia Eucalypti viminalis и Camellia sinensis, в отношении P. aeruginosa наилучшие показатели синергического взаимодействия наблюдались у Folia Eucalypti viminalis и Allium sativum.
Присутствие в составе выбранных лекарственных растений вторичных метаболитов, обладающих антибактериальным действием, таких как рутин, галловая кислота, тимол, танин, объясняют природу возникновения антибактериального эффекта (действие антибиотика в эксперименте не учитывалось, так как концентрации брались ниже МИК этого антибиотика) [1].

Чтобы объяснить наблюдаемый синергетический эффект, от добавления в среду вторичных метаболитов при культивировании, использовались методы квантовой химии. Рассматривался вариант возможности образования в среде межмолекулярных или молекулярных комплексов между вторичными метаболитами фенольного происхождения и фосфомицином и цефтазидимом. Доказательная база строилась по распределению электронной плотности на атомах и электростатическому потенциалу. Области, имеющие большой отрицательный потенциал, сопряженный с наличием гидроксильных или карбонильных групп, должны стать электрофильными центрами, а молекула или ее участок в целом нуклеофилом. Все это должно привести к образованию, за счет электростатических сил, межмолекулярного комплекса между молекулами вторичных метаболитов и молекулами антибиотиков, а в случае наличия комплексообразователя к смешаннолигандному комплексу. На рис. 1 представлены молекулярно-электростатический потенциал (МЭП) и распределение электронной плотности вторичных метаболитов и антибиотиков.

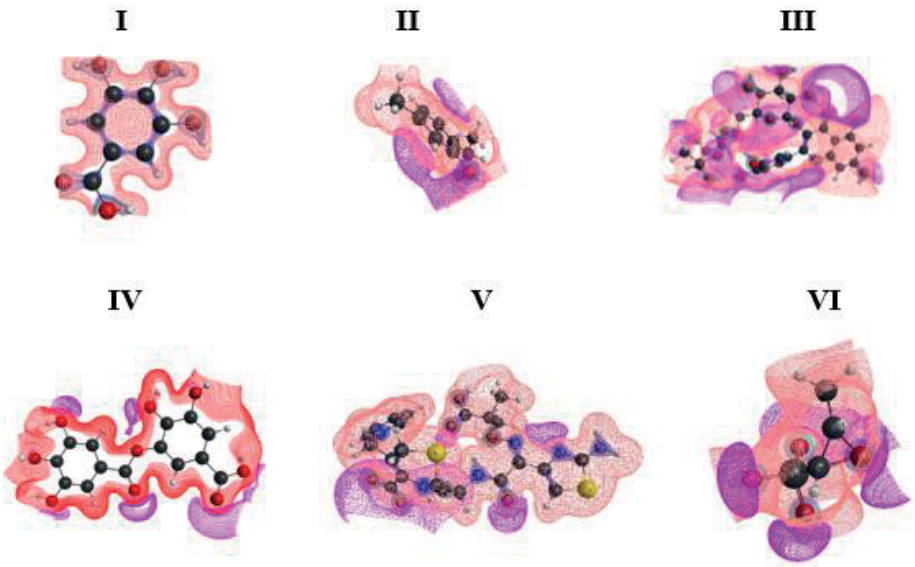


Рис. 1. Молекулярно-электростатический потенциал и распределение электронной плотности (I – gallic acid, II – thymol, III – routine, IV – tannin, V – ceftazidime, VI – fosfomycin)

По полученным данным МЭП исследуемых структур было высказано предположение, что все соединения, за исключением галловой кислоты, могут выступать как электро- так и нуклеофилами. Наибольшая электронная плотность сосредоточена на атомах кислорода и азота, в некоторых случаях распределена по ароматическому кольцу. В основном отрицательные области электростатического потенциала расположены вдоль гидроксильных групп.

В случае с противомикробными средствами для фосфомицина электрофильные участки расположены точечно над гидроксильными группами, для цефтазидима форма электрофильных участков сложная.

На рис. 2 показаны центры, к которым может присоединиться противоположная молекула вторичного метаболита или антибиотиков. Это может быть выражено в маскировании молекул АБ от бета-лактамаз и как следствие, в уменьшении их необходимой концентрации.

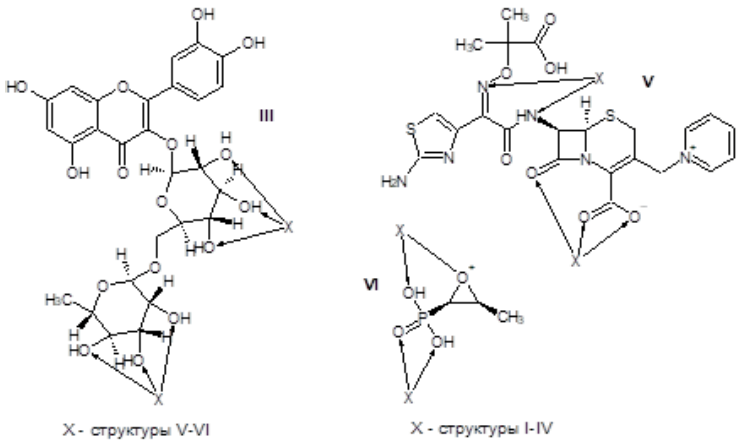


Рис. 2. Возможные конфигурации межмолекулярных комплексов вторичных метаболитов с молекулами фосфомицина и цефтазидима

Основная задача спектрального анализа заключалась в проверке, вступают ли антибиотики с молекулами вторичных метаболитов в химическую реакцию или с другими молекулами в экстрактах растений. По полученным спектрам было выявлено, что не вступают.

Ключевые частоты вторичных метаболитов, находились на одинаковых полосах и в спектрах экстрактов хорошо просматривались. Полосы поглощения в противомикробных средствах почти совпали с полосами поглощения у вторичных метаболитов, однако, в экстрактах, при совместном введении в среду, не заметны на фоне. Это говорит о малой концентрации антибиотиков, но также указывает и на то, что в химическое взаимодействие цефтазидим и фосфомицин не вступают, так как отсутствуют дополнительные пики, свидетельствующие о продуктах реакции. Следовательно, антибиотики в растворе могут образовывать только межмолекулярные комплексы [2].

Выводы: В результате проведенных исследований было установлено, что антибактериальная активность может быть обусловлена как возможным синергетическим эффектом между взаимодействующими комплексами вторичных метаболитов и антибиотиков, так и индивидуальными механизмами, присущими конкретному действующему агенту, присутствующему в экстракте. При этом, антибактериальный эффект, показанный комплексным взаимодействием антибиотиков и растительных экстрактов с фенолсодержащими соединениями, будет зависит от внешних условий взаимодействия (среда, температура, давление, влажность и т.д., так и внутренних факторов (изменение конформаций взаимодействия, условий протекания реакций и пр.).

Chemical modification of antimicrobial complexes with organic molecules

Plotnikova Yu.*, Barysheva E., Bibartseva E., Baranova O.
Orenburg State University, Orenburg, Russia
** shik8mail@mail.ru*

Key words: gallic acid; molecular electrostatic potential; quantum chemical modelling; synergetic interaction

Purpose: The purpose of the study was to determine the synergistic antibacterial effect of the combined use of antibiotics and secondary metabolites of phenolic origin, as well as to determine the nature of their interaction.
Materials and methods: The research material was clinical isolates of P.aeruginosa and S. aureus. The pre-extracted phenolic components of 10 medicinal plants were studied: Herba Achilleae millefolii; Folia Eucalypti viminalis; Urticae folia; Sortex Quercus; Nerba; Camellia; Ribes nigrum L.; Sorbus aucuparia L.; Allium sativum; Dianthus in combination with the antibiotic drugs ceftazidime and fosfomycin.
The combined action of antibiotics and phenol-containing extracts was tested using the agar wells method.
During the quantum chemical calculation, the molecules of the studied substances were first optimized in a small approximation of PBE/3-21+G. A configuration search for the optimal spatial arrangement of functional groups was carried out, and at the next stage, the calculation of the studied molecules was performed, with the construction of maps of the molecular electrostatic potential (MEP). The molecules calculated at the first stage, taking into account the MEP, were paired and optimized based on the density functional theory in the PBE/3-21+G approximation.
The spectra of pure substances and putative intermolecular complexes were obtained using the Fourier spectrometer INFRALUM FT–02 and the Fourier infrared spectrometer Bruker Alpha (without the prefix NPVO).
Results: Initially, studies of antimicrobial activity were conducted in vitro, with the combined introduction of secondary metabolites of phenolic origin and antibiotics (fosfomycin and ceftazidime) into the medium. With respect to S.aureus, the most pronounced synergistic effect was noted for Folia Eucalypti viminalis and Camellia sinensis, with respect to P.aeruginosa, the best indicators of synergistic interaction were observed in Folia Eucalypti viminalis and Allium sativum.
The presence of secondary metabolites with antibacterial action in the composition of selected medicinal plants, such as rutin, gallic acid, thymol, tannin, explain the nature of the antibacterial effect (the effect of the antibiotic was not taken into account in the experiment, since concentrations were taken below the MIC of this antibiotic) [1].
Quantum chemistry methods were used to explain the observed synergistic effect of adding secondary metabolites to the medium during cultivation. The possibility of education in the environment was considered
intermolecular or molecular complexes between secondary metabolites of phenolic origin and fosfomycin and ceftazidime. The evidence base was based on the distribution of electron density on atoms and electrostatic potential. Regions with a large negative potential associated with the presence of hydroxyl or carbonyl groups should become

electrophilic centers, and the molecule or its site as a whole should become a nucleophile. All this should lead to the formation, due to electrostatic forces, of an intermolecular complex between molecules of secondary metabolites and molecules of antibiotics, and in the case of a complexing agent to a mixed ligand complex. Figure 1 shows the molecular electrostatic potential (MEP) and the electron density distribution of secondary metabolites and antibiotics

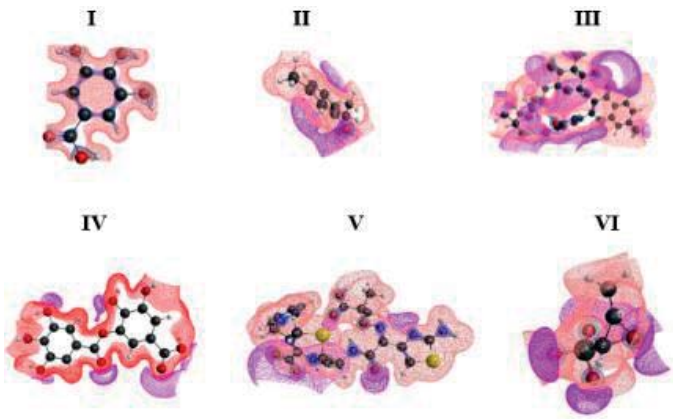


Fig. 1. Molecular electrostatic potential and electron density distribution (I – gallic acid, II – thymol, III – routine, IV – tannin, V – ceftazidime, VI – fosfomycin)

According to the obtained MEP data of the studied structures, it was suggested that all compounds, with the exception of gallic acid, can act as both electro and nucleophiles. The highest electron density is concentrated on oxygen and nitrogen atoms, in some cases distributed along the aromatic ring. Basically, the negative regions of the electrostatic potential are located along the hydroxyl groups. In the case of antimicrobials for fosfomycin, the electrophilic sites are located pointwise above the hydroxyl groups, for ceftazidime, the shape of the electrophilic sites is complex. Figure 2 shows the centers to which the opposite molecule of a secondary metabolite or antibiotics can join. This can be expressed in masking the AB molecules from beta-lactamases and, as a result, in reducing their required concentration.

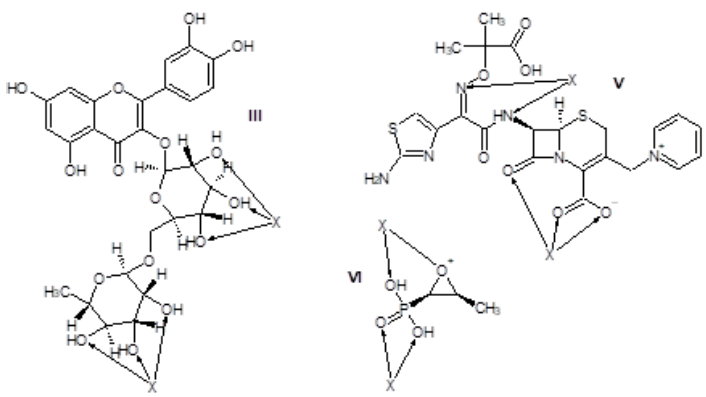


Fig. 2. Possible configurations of intermolecular complexes of secondary metabolites with phosphomycin and ceftazidime molecules

The main task of spectral analysis was to check whether antibiotics react chemically with molecules of secondary metabolites or with other molecules in plant extracts. According to the obtained spectra, it was revealed that they do not enter.

The key frequencies of the secondary metabolites were on the same bands and were clearly visible in the spectra of the extracts. The absorption bands in antimicrobials almost coincided with the absorption bands in secondary metabolites, however, in extracts, when co-administered into the medium, they are not noticeable against the background. This indicates a low concentration of antibiotics, but also indicates that ceftazidime and fosfomycin do not enter into chemical interaction, since there are no additional peaks indicating reaction products. Consequently, antibiotics in solution can only form intermolecular complexes [2].

Conclusions: As a result of the conducted studies, it was found that the antibacterial activity may be due to both a possible synergistic effect between interacting complexes of secondary metabolites and antibiotics, and individual mechanisms inherent in a specific active agent present in the extract. At the same time, the antibacterial effect shown by the complex interaction of antibiotics and plant extracts with phenol-containing compounds will depend on the external conditions of interaction (environment, temperature, pressure, humidity, etc., and internal factors (changes in interaction conformations, reaction conditions, etc.).

Список литературы/References

1. Плотникова Ю.А., Барышева Е.С. Биохимические аспекты комплексного антибактериального действия фенолсодержащих вторичных метаболитов лекарственных растений в сочетании с антибиотиками. В: Фундаментальные исследования в области химии, биологии и экологии. Оренбург: Оренбургский государственный университет, 2022;149-151
2. Плотникова Ю.А., Барышева Е.С., Пешков С.А. Химическая модификация противомикробных комплексов органическими молекулами с целью получения смешанно-лигантных антирезистентных агентов. *Технологии живых систем*. 2023;20(2):42-52. doi 10.18127/j20700997-202302-05

doi 10.18699/bgrs2024-3.2-12

Синтез и изучение фотолиза при облучении светом зеленой области нитрозо производных BODIPY

Рыкунов Д.А.[1]*, Карогодина Т.Ю.[1, 2], Воробьев А.Ю.[1]

[1] *Новосибирский институт органической химии СО РАН, Новосибирск, Россия*

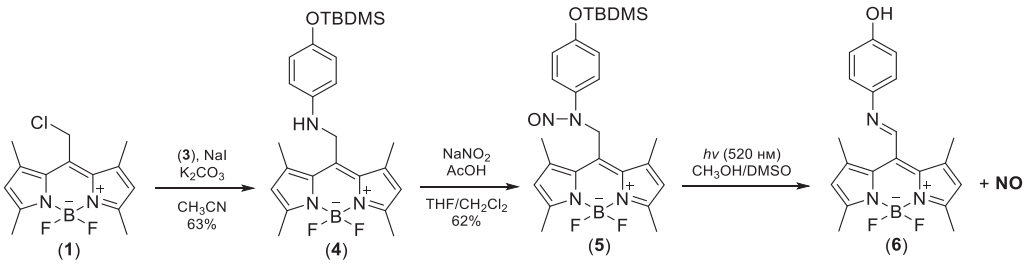
[2] *Новосибирский государственный университет, Новосибирск, Россия*

** d.rykunov@g.nsu.ru*

Ключевые слова: фотодоноры NO; BODIPY; фотоактивация светом видимой области

Мотивация и цель: Одним из направлений в области фотохимии является разработка фотоактивируемых доноров оксида азота (NO) – важной сигнальной молекулы, которая участвует в регуляции множества физиологических процессов. Нарушение регуляции концентрации NO часто приводит к различным патологическим состояниям. Высокая же концентрация NO может приводить к апоптозу и гибели клеток, что представляет интерес для борьбы с опухолями. Однако достижение высоких локальных концентраций NO с помощью классических доноров оксида азота затруднительно, в связи с чем разработка фотогенератов NO стала актуальной задачей. Световое излучение позволяет малоинвазивно реализовать пространственно-временной и концентрационный контроль. Актуальным направлением является разработка фотодоноров NO, активация которых происходит при облучении видимым и ближним ИК-светом, что может представлять интерес для фотодинамической терапии опухолей. BODIPY и его производные являются перспективными молекулами в дизайне фотоактивируемых доноров NO за счет узких полос поглощения и высокой биосовместимости.

Результаты: разработан синтез *N*-(4-(трет-бутилдиметилсилил)-фенил)-*N*-(2-(1,3,5,7-тетраметил-BODIPY-10-ил)-метил)нитрозо амида (5) на основе производного BODIPY (1) и проведено изучение фотолиза амида (5) при облучении зеленым светом (520 нм) с образованием имина (6) и выделением NO.



Выводы: Показано, что *мезо*-производные BODIPY могут выступать эффективными фотогенераторами NO, активация которых происходит при облучении светом видимой области.

Synthesis and study of photolysis upon irradiation with light from the green region of nitroso derivatives BODIPY

Rykunov D.A.[1]*, Vorob'ev A.Y.[1], Karogodina T.Y.[1, 2]

[1] *N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia*

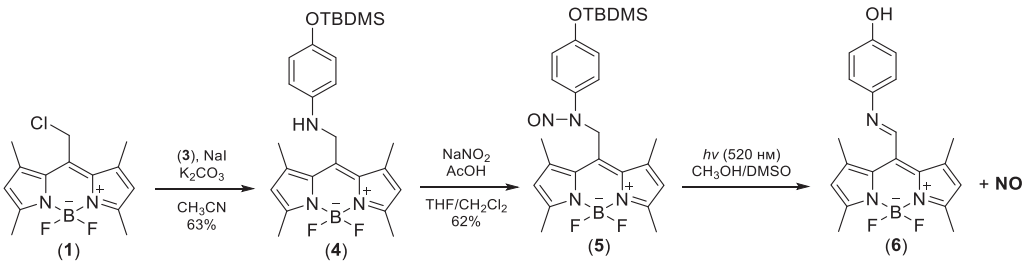
[2] *Novosibirsk State University, Novosibirsk, Russia*

** d.rykunov@g.nsu.ru*

Key words: photodonors NO; BODIPY; photoactivation by visible light

Motivation and Aim: One direction in the field of photochemistry is the development of photoactivatable NO donors, an important signaling molecule that is involved, for example, in the regulation of smooth muscle tone and in the response of the immune system. A high concentration of NO can lead to tissue damage or various pathological processes, and therefore the development of NO photogenerators is relevant due to the possibility of spatiotemporal and concentration control. A current direction is the development of NO photodonors, the activation of which occurs upon irradiation with visible and near-IR light, which may be of interest for photodynamic therapy of tumors. BODIPY and its derivatives are promising molecules in the design of photoactivatable NO donors due to their narrow absorption bands and high biocompatibility.

Results: A synthesis of *N*-(4-(tert-butyldimethylsilyl)-phenyl)-*N*-(2-(1,3,5,7-tetramethyl-BODIPY-10-yl)-methyl)nitroso amide (5) was developed based on the derivative BODIPY (1) and studied the photolysis of amide (5) under irradiation with green light (520 nm) with the formation of imine (6) and the release of NO.



Conclusion: It was shown that *meso*-derivatives of BODIPY can act as effective photogenerators of NO, the activation of which occurs when irradiated with light in the visible region.

doi 10.18699/bgrs2024-3.2-14

Исследование ингибиторов Tdp1 на клетках нокаутных по генам белков системы репарации ДНК PARP1 и Tdp1

Чепанова А.А.[1]*, Чернышова И.А.[1], Корниенко Т.Е.[1], Захаренко А.Л.[1], Дырхеева Н.С.[1], Дреничев М.С.[2], Филимонов А.С.[3], Лузина О.А.[3], Салахутдинов Н.Ф.[3], Лаврик О.И.[1]

[1] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

[2] *Институт молекулярной биологии им. В.А. Энгельгарда РАН, Москва, Россия*

[3] *Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия*

* *arinachepanova@mail.ru*

Ключевые слова: система репарации ДНК; тирозил-ДНК-фосфодиэстераза 1; топоизомераза 1; противоопухолевые препараты; ингибиторы

Мотивация и цель: На сегодняшний день известно, что фермент репарации тирозил-ДНК-фосфодиэстеразы 1 (Tdp1) играет ключевую роль в удалении повреждений ДНК, образующихся при ингибировании топоизомеразы 1 (Top1) клинически используемыми противоопухолевыми препаратами (иринотекан, топотекан), а также участвует в репарации повреждений ДНК, вызванных другими противоопухолевыми препаратами [1, 2]. Также имеются данные о взаимодействии Tdp1 с другим ферментом системы репарации, а именно поли- (АДФ-рибоза)-полимеразой 1 (PARP1). PAR-илирование (посттрансляционная модификация белков, катализируемая PARP1) регулирует ферментативную активность Tdp1, стабилизирует фермент и стимулирует рекрутирование Tdp1 в места повреждений ДНК, вызванные комплексами Top1cc [3, 4]. Основываясь на этих данных, мы считаем, что необходимо изучать действие ингибиторов Tdp1 на клетки с учетом роли PARP1. Для получения новой информации о работе ингибиторов Tdp1 мы использовали клеточные линии, нокаутные по генам Tdp1 (Tdp1-/-) или PARP1 (PARP1-/-) [5, 6]. Для изучения нами были выбраны два соединения различной природы из ранее найденных, оказавших выраженное сенсибилизирующее действие на опухолевые клетки как *in vitro*, так и *in vivo*. Одно из соединений является бидериватизированным производным усниновой кислоты (лабораторный шифр AF-185, синтезировано в НИОХ СО РАН) [7], другое (лабораторный шифр 577, синтезировано в ИМБ РАН) [8] – липофильным производным пуринового нуклеозида.

Методы и алгоритмы: Влияние ингибиторов Tdp1 на выживаемость клеток и на цитотоксический эффект топотекана было изучено с помощью МТТ-теста путем колориметрического измерения количества формазана, конвертированного из 3-(4,5-диметилтиазол-2-ил)-2,5-дифенил-2Н-тетразолия бромида (МТТ) клетками, подвергшимися воздействию соединений (количество выживших клеток). Клетки культивировали в среде DMEM/F12, содержащей 10 % FBS и 1 % раствор антибиотиков и антимикотиков (100 ед./мл пенициллина, 0.1 мкг/мл стрептомицина, 0.25 мкг/мл амфотерицина), в атмосфере 5 % CO_2 при 37 °C. Клетки пересеивали раз в 3–4 дня для поддержания экспоненциального роста.

Результаты: Была изучена собственная цитотоксичность/антипролиферативная активность ингибиторов Tdp1 на контрольной клеточной линии HEK293A

(получена из эмбриональных почек человека) дикого типа (WT), трех клонах Tdp1-/- и клоне PARP1-/- с использованием МТТ-теста. Показано, что соединение AF-185 малотоксично как для клеток HEK293A WT, так и для нокаутных по генам белков репарации ДНК PARP1 и Tdp1 (не менее 80 % живых клеток). Другой ингибитор Tdp1, соединение 577, оказывает более цитотоксичное/антипролиферативное действие на клетки, примерно сопоставимое для HEK293A WT и всех нокаутных клонов, что говорит об отсутствии связи токсического эффекта с наличием или отсутствием Tdp1 в клетках и о вероятных побочных мишенях этого соединения. Также изучено влияние ингибиторов Tdp1 на цитотоксический эффект топотекана на контрольных клетках HEK293A WT и клетках Tdp1-/- и PARP1-/-. Несмотря на высокую ингибирующую активность выбранных производных в отношении Tdp1 в экспериментах *in vitro*, на клетках неопухолевого происхождения эти соединения не продемонстрировали усиления цитотоксического эффекта топотекана.

В связи с отсутствием сенсибилизирующего эффекта на клетках неракового типа, мы изучили влияние выбранных соединений на выживаемость контрольной клеточной линии аденокарциномы альвеолярных базальных эпителиальных клеток человека A549 и клеток A549 с нокаутом по гену *Tdp1* (клеточные клоны B5, B6, B10) в монорежиме и в комбинации с топотеканом. Соединение AF-185 оказалось малотоксичным как для A549 WT, так и для A549 Tdp1-/-. Соединение 577 оказало более цитотоксичное/антипролиферативное действие на клетки, примерно сопоставимое для A549 WT и A549 Tdp1-/-, как и в случае клеток неракового происхождения HEK293A.

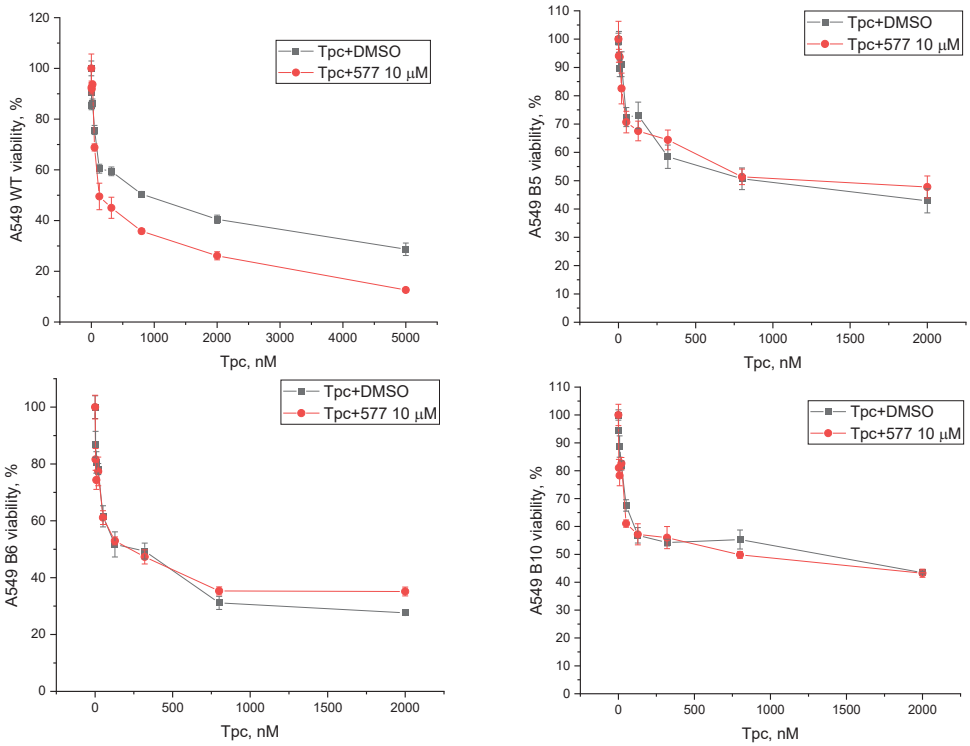


Рис. 1. Дозозависимое влияние топотекана в комбинации с липофильным производным пуринового нуклеозида 577 в концентрации 10 мкМ на выживаемость контрольной клеточной линии A549 WT и клонов Tdp1-/-

Далее было изучено влияние топотекана на выживаемость контрольной клеточной линии и выживаемость клонов A549 Tdp1-/-. Показано, что нокаутные по гену *Tdp1* раковые клеточные линии (B5, B6, B10) более чувствительны к топотекану, по сравнению с клетками дикого типа, как и в случае клеток неракового происхождения HEK293A. Также изучено влияние ингибиторов Tdp1 на цитотоксический эффект топотекана на контрольных клетках A549 WT и на клетках Tdp1-/- (рис. 1, графики приведены только для соединения 577). Показано, что усиление действия топотекана в присутствии ингибиторов Tdp1 наблюдалось только для клеточной линии дикого типа A549 WT и полностью отсутствовало для мутантных клеток.

Выводы: Усиление действия топотекана в присутствии ингибиторов Tdp1 наблюдалось только для клеточной линии дикого типа A549 WT и полностью отсутствовало для мутантных клеток. Это может говорить о том, что синергетическое действие топотекана совместно с ингибиторами Tdp1 обусловлено именно подавлением активности Tdp1 выбранными ингибиторами.

Финансирование: Исследование поддержано грантом РНФ № 23-74-01078.

Study of Tdp1 inhibitors on cells knockout for the genes of the DNA repair system proteins PARP1 and Tdp1

Chepanova A.A.[1]*, Chernyshova I.A.[1], Kornienko T.E.[1], Zakharenko A.L.[1], Dyrkheeva N.S.[1], Drenichev M.S.[2], Filimonov A.S.[3], Luzina O.A.[3], Salakhutdinov N.F.[3], Lavrik O.I.[1]

[1] *Novosibirsk Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

[3] *N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia*

* *arinachepanova@mail.ru*

Key words: DNA repair system; tyrosyl-DNA phosphodiesterase 1; topoisomerase 1; antitumor drugs; inhibitors

Motivation and Aim: Today it is known that the repair enzyme tyrosyl-DNA phosphodiesterase 1 (Tdp1) plays a key role in the removal of DNA damage formed when topoisomerase 1 (Top1) is inhibited by clinically used anticancer drugs (irinotecan, topotecan), and is also involved in the repair of DNA damage caused by other antitumor drugs [1, 2]. There is also evidence of the interaction of Tdp1 with another enzyme of the repair system, namely poly(ADP-ribose) polymerase 1 (PARP1). PARylation (post-translational modification of proteins catalyzed by PARP1) regulates the enzymatic activity of Tdp1, stabilizes the enzyme and stimulates the recruitment of Tdp1 to sites of DNA damage caused by Top1cc complexes [3, 4]. Based on these data, we believe that it is necessary to study the effect of Tdp1 inhibitors on cells, taking into account the role of PARP1. To obtain new information about the work of Tdp1 inhibitors, we used cell lines knockout for the Tdp1 (Tdp1-/-) or PARP1 (PARP1-/-) genes [5, 6]. For study, we selected two compounds of different nature from those previously found, which had a pronounced sensitizing effect on tumor cells both in vitro and in vivo. One of the compounds is a biderivatized derivative of usnic acid (laboratory code AF-185, synthesized at the Institute of Organic Chemistry of the Siberian Branch of the Russian

Academy of Sciences) [7], the other compound (laboratory code 577, synthesized at the Institute of Biochemistry of the Russian Academy of Sciences) [8] is a lipophilic derivative of purine nucleoside.

Methods and Algorithms: The effect of Tdp1 inhibitors on cell survival and on the cytotoxic effect of topotecan was studied using the MTT test by colorimetric measurement of the amount of formazan converted from 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) cells exposed to compounds (number of surviving cells). Cells were cultured in DMEM/F12 medium containing 10 % FBS and 1 % solution of antibiotics and antimycotics (100 units/ml penicillin, 0.1 μg/ml streptomycin, 0.25 μg/ml amphotericin), in an atmosphere of 5 % CO_2 at 37 °C. Cells were subcultured every 3–4 days to maintain exponential growth.

Results: The intrinsic cytotoxicity/antiproliferative activity of Tdp1 inhibitors was studied in the wild-type (WT) control cell line HEK293A (derived from human embryonic kidney), three Tdp1-/- clones, and a PARP1-/- clone using the MTT assay. It has been shown that the AF-185 compound has low toxicity both for HEK293A WT cells and for DNA repair protein genes knockout PARP1 and Tdp1 (at least 80 % of living cells). Another Tdp1 inhibitor, compound 577, has a more cytotoxic/antiproliferative effect on cells, approximately comparable for HEK293A WT and all knockout clones, which indicates that the toxic effect is not related to the presence or absence of Tdp1 in cells and the likely side targets of this compound.

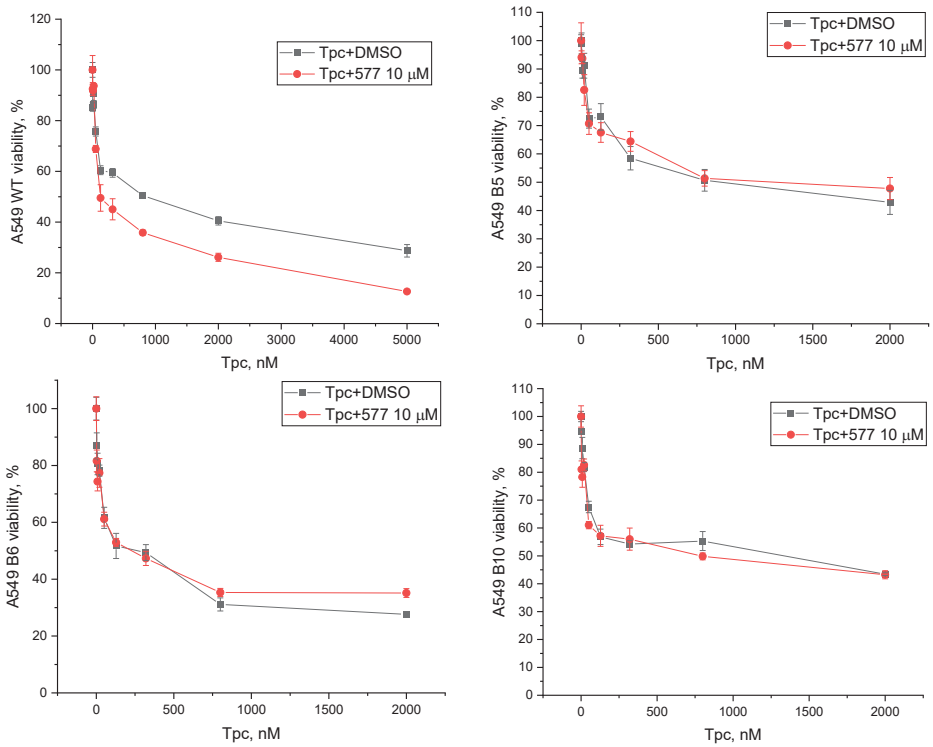


Fig. 1. Dose-dependent effect of topotecan in combination with the lipophilic derivative of purine nucleoside 577 at a concentration of 10 μM on the survival of the control cell line A549 WT and Tdp1-/- clones

The effect of Tdp1 inhibitors on the cytotoxic effect of topotecan was also studied in control HEK293A WT cells and Tdp1-/- and PARP1-/- cells. Despite the high inhibitory activity of the selected derivatives against Tdp1 in in vitro experiments, in cells of non-tumor origin these compounds did not demonstrate an increase in the cytotoxic effect of topotecan. Due to the lack of a sensitizing effect on non-cancerous cells, we studied the effect of the selected compounds on the survival of the control cell line adenocarcinoma of human alveolar basal epithelial cells A549 and A549 cells with a knockout of the Tdp1 gene (cell clones B5, B6, B10) in mono mode and in combination with topotecan. Compound AF-185 was found to have low toxicity for both A549 WT and A549 Tdp1-/-. Compound 577 had a more cytotoxic/antiproliferative effect on cells, approximately comparable for A549 WT and A549 Tdp1-/-, as in the case of non-cancerous HEK293A cells.

Next, the effect of topotecan on the survival of the control cell line and the survival of A549 Tdp1-/- clones was studied. It has been shown that cancer cell lines knockout for the Tdp1 gene (B5, B6, B10) are more sensitive to topotecan compared to wild-type cells, as is the case with non-cancerous HEK293A cells. The effect of Tdp1 inhibitors on the cytotoxic effect of topotecan was also studied on control A549 WT cells, as well as on Tdp1-/- cells (Fig. 1, graphs are shown only for compound 577). It was shown that the enhanced effect of topotecan in the presence of Tdp1 inhibitors was observed only for the wild type A549 WT cell line and was completely absent for mutant cells.

Conclusion: An increase in the effect of topotecan in the presence of Tdp1 inhibitors was observed only for the wild-type cell line A549 WT and was completely absent for mutant cells, this may indicate that the synergistic effect of topotecan together with Tdp1 inhibitors is due precisely to the suppression of Tdp1 activity by the selected inhibitors.

Funding: The study is supported by the Russian Scientific Foundation grant No. 23-74-01078.

Список литературы/References

1. Brettrager E.J., van Waardenburg R.C.A.M. Targeting Tyrosyl-DNA phosphodiesterase I to enhance toxicity of phosphodiester linked DNA-adducts. *Cancer Drug Resist*. 2019;2(4):1153-1163. doi 10.20517/cdr.2019.91
2. Pommier Y. et al. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol*. 2010;17(5):421-433. doi 10.1016/j.chembiol.2010.04.012
3. Brettrager E.J. et al. Tyrosyl-DNA Phosphodiesterase I N-Terminal Domain Modifications and Interactions Regulate Cellular Function. *Genes*. 2019;10(11):897. doi 10.3390/genes10110897
4. Sun Y. et al. Debulking of topoisomerase DNA-protein crosslinks (TOP-DPC) by the proteasome, non-proteasomal and non-proteolytic pathways. *DNA Repair*. 2020;94:102926. doi 10.1016/j.dnarep.2020.102926
5. Dyrkheeva N.S. et al. New Hybrid Compounds Combining Fragments of Usnic Acid and Thioether Are Inhibitors of Human Enzymes TDP1, TDP2 and PARP1. *Int J Mol Sci*. 2021;22(21):11336. doi 10.3390/ijms222111336
6. Dyrkheeva N.S. et al. New Hybrid Compounds Combining Fragments of Usnic Acid and Monoterpenoids for Effective Tyrosyl-DNA Phosphodiesterase 1 Inhibition. *Biomolecules*. 2021;11(7):973. doi 10.3390/biom11070973
7. Kornienko T.E. et al. Enhancement of the Antitumor and Antimetastatic Effect of Topotecan and Normalization of Blood Counts in Mice with Lewis Carcinoma by Tdp1 Inhibitors-New Usnic Acid Derivatives. *Int J Mol Sci*. 2024;25(2):1210. doi 10.3390/ijms25021210
8. Chernyshova I.A. et al. The Lipophilic Purine Nucleoside-Tdp1 Inhibitor-Enhances DNA Damage Induced by Topotecan In Vitro and Potentiates the Antitumor Effect of Topotecan In Vivo. *Molecules*. 2022;28(1):323. doi 10.3390/molecules28010323

doi 10.18699/bgrs2024-3.2-15

Основные принципы аллостерической регуляции гормональных рецепторов, сопряженных с G-белками

Шпаков А.
Институт эволюционной физиологии и биохимии им. И.М. Сеченова РАН, Санкт-Петербург, Россия
alex_shpakov@list.ru

Ключевые слова: аллостерическая регуляция; G-белок-сопряженный рецептор; сигнальная трансдукция; трансмембранный домен; ортостерический агонист; рецептор лютеинизирующего гормона; пепдуцин; инверсионный агонист; аутоантитела к рецепторам

Рецепторы, сопряженные с αβγ-гетеротримерными G-белками (GPCR), играют ключевую роль в передаче гормональных сигналов к внутриклеточным сигнальным каскадам и потому ответственны за регуляцию экспрессии генов и посттрансляционную модификацию белков, вовлечены в контроль пролиферации, апоптоза, дифференцировки. Для каждого GPCR имеется свой эндогенный регулятор, как правило, гормон или нейромедиатор (ортостерический агонист), который с высоким сродством связывается с лиганд-связывающим ортостерическим сайтом GPCR, переводя рецептор в активное состояние и запуская тем самым процесс сигнальной трансдукции, опосредуемый через различные типы G-белков и β-аррестинов. На протяжении многих лет основные исследования в области фармакологии GPCR были направлены на создание синтетических лигандов ортостерического сайта с определенным спектром биологической активности, вследствие чего значительная часть лекарств, мишенями которых являются GPCR, – это лиганды их ортостерического сайта. В последнее время акцент в изучении GPCR сместился в сторону исследования их аллостерической регуляции, и основной импульс этому придали исследования в области расшифровки пространственной структуры GPCR, стартовавшие в конце 2000-х гг., после открытия 3D структуры β_2-адренергического рецептора [1]. Сначала считали, что в молекуле GPCR имеется сравнительно мало аллостерических сайтов, от одного до 3–4. Предполагали также, что ортостерический и аллостерический сайты должны быть локализованы в разных субдоменах GPCR и не пересекаться. Однако в дальнейшем было показано, что число аллостерических сайтов в GPCR существенно больше, локализованы они во всех основных субдоменах рецептора и могут физически перекрываться с ортостерическим сайтом [2].

Наиболее характерной локализацией аллостерических сайтов являются внеклеточные субдомены GPCR, в том числе внешний вестибюль трансмембранного (ТМ) канала, формируемый внеклеточными петлями и внешними окончаниями гидрофобных ТМ спиралей [2, 3]. Функции этих субдоменов состоят в первичном опознавании и низкоаффинном связывании гормонов и формировании гомо- и гетеродимерных рецепторных комплексов [4]. Кроме того, в некоторых GPCR, например в рецепторах лютеинизирующего (ЛГ) и тиреотропного (ТТГ) гормонов, имеются значительные по размеру эктодомены, которые содержат высокоаффинные ортостерические сайты и при этом являются мишенями для

аллостерических регуляторов, которые не только модулируют связывающие характеристики GPCR, но и способны сами регулировать их активность, как это показано для аутоантител к рецептору ТТГ [5, 6]. Эндогенными регуляторами внеклеточных аллостерических сайтов, наряду с аутоантителами, могут быть белки, модифицирующие активность рецепторов (RAMP), протомеры других рецепторов, в том числе рецепторы тирозинкиназного типа и ионотропные рецепторы, а также пептидные регуляторы, например агути-подобный пептид в случае М3- и М4-меланокортиновых рецепторов. В последние годы разработаны низкомолекулярные соединения, способные связываться с полостью внешнего вестибюля ТМ канала [3]. Они влияют на доступность ортостерического сайта для агонистов, а также способны регулировать активность лиганд-свободного GPCR.

Другой локализацией аллостерических сайтов является внутренняя полость ТМ канала, которая доступна для сравнительно небольших молекул и ионов. Внутри нее обычно располагаются несколько аллостерических сайтов, которые могут перекрываться между собой и с ортостерическим сайтом. Эндогенными лигандами аллостерических сайтов, расположенных в ТМ канале, чаще всего являются ионы натрия и некоторых двухвалентных металлов (Zn^{2+}, Mg^{2+}, Ca^{2+}, Cu^{2+}) [3, 7]. В настоящее время разработано немало низкомолекулярных соединений, которые способны проникать внутрь ТМ канала и связываться с расположенными в нем аллостерическими сайтами. Так, для регуляции рецепторов ЛГ и ТТГ нами были разработаны производные тиено[2,3-d]-пиримидина с активностью как полных и инверсионных агонистов, так и нейтральных антагонистов этих рецепторов [3, 8, 9]. При введении крысам они были активны и при внутрибрюшинном, и при пероральном способе доставки, что указывает на их стабильность и хорошую всасываемость в желудочно-кишечном тракте. Низкомолекулярные аллостерические агонисты не конкурировали с гонадотропинами и ТТГ за связывание с рецептором, тем самым не препятствуя регуляции эндокринных функций эндогенными гормонами.

В ТМ домене аллостерические сайты также могут располагаться на его боковых поверхностях, контактируя с липидной фазой мембраны [2, 7]. Эндогенными лигандами таких сайтов могут быть липиды плазматической мембраны, в первую очередь холестерин и фосфолипиды, а также некоторые стероидные гормоны и их прекурсоры (прогестерон, прегненолон и др.). Взаимодействие липидов с этими аллостерическими сайтами GPCR обеспечивает тесные взаимосвязи между активностью рецепторов, с одной стороны, и липидным составом и физико-химическими свойствами мембраны, с другой [3, 7].

Еще одной локализацией аллостерических сайтов являются цитоплазматические петли и ориентированный вовнутрь клетки вестибюль ТМ канала. Эндогенными регуляторами этих сайтов служат гетеротримерные G-белки и β-аррестины, которые не только функционируют, как трансдукторы гормонального сигнала, но и осуществляют аллостерическую модуляцию активности рецепторов, определяя их аффинность к ортостерическому агонисту и предвзятость внутриклеточного сигналинга [3]. В последние годы ведется интенсивная разработка синтетических аллостерических регуляторов внутриклеточных сайтов – как низкомолекулярных соединений, так и модифицированных гидрофобными радикалами пептидов, производных функционально важных внутриклеточных участков GPCR, называемых пепдуцинами [10]. Нами разработаны модифицированные пальмитатом

пептуцины NKDTKIAKK-Nle-A(562–572)-K(Palm)A и 612–627(Palm), соответствующие участкам третьей цитоплазматической петли рецепторов ЛГ и ТТГ, которые *in vitro* и *in vivo* проявляли активность частичных агонистов тех рецепторов, производными первичной структуры которых они являлись [3]. В основе действия пептуцинов лежит изменение конформации внутриклеточных аллостерических сайтов вследствие их специфичного взаимодействия с комплементарными участками цитоплазматических петель рецептора и внутриклеточным вестибюлем ТМ канала.

Наряду с множественностью аллостерических сайтов необходимо принимать во внимание широкий спектр фармакологической активности их лигандов. Эти лиганды могут повышать или снижать сродство ортостерического агониста к рецептору и эффективность его действия, будучи соответственно позитивными (PAM) или негативными (NAM) аллостерическими модуляторами [2, 7]. Они могут тонко настраивать сигнальную трансдукцию, влияя на предвзятость действия ортостерического агониста на внутримолекулярные мишени, и тогда их относят к классу «молчащих» аллостерических модуляторов (SAM). Они могут быть наделены собственной активностью полных/частичных агонистов, инверсионных агонистов и нейтральных антагонистов или сочетать в себе свойства аллостерических агонистов и модуляторов (Ago-PAM, Ago-NAM).

Выводы: Таким образом, в основе аллостерической регуляции GPCR лежит множественность аллостерических сайтов, различающихся по локализации, структурно-функциональной организации и эндогенным регуляторам, причем эти сайты способны не только влиять на конформацию и связывающие характеристики ортостерического сайта, но и взаимодействовать между собой. Наряду с такими эндогенными лигандами аллостерических сайтов GPCR, как простые ионы, липиды, аминокислоты, полипептиды, важную роль в аллостерической регуляции GPCR играют процессы образования их комплексов с G-белками, β-аррестинами, RAMP, другими адаптерными и регуляторными белками, а также формирование гомо- и гетероди(олиго)мерных рецепторных комплексов. Все это существенно усложняет и обогащает механизмы GPCR-опосредуемой сигнальной трансдукции и создает мощную теоретическую и практическую базу для создания большого числа различающихся по своей селективности и фармакологическому профилю аллостерических регуляторов. Следует отметить, что такие регуляторы, в отличие от ортостерических агонистов, действуют более мягко, что снижает риски десенситизации GPCR, а также более специфичны по отношению к внутриклеточным эффекторам, что позволяет избежать ряда побочных эффектов, характерных для гормональной терапии. Для некоторых типов GPCR, например для рецептора ТТГ, для которых отсутствуют подходящие фармакологические регуляторы ортостерического сайта, создание аллостерических регуляторов является безальтернативным подходом [3].

Финансирование: Исследование поддержано грантом Российского научного фонда (проект № 19-75-20122).

Basic principles of allosteric regulation of hormonal receptors coupled to G proteins

Shpakov A.
Sechenov Institute of Evolutionary Physiology and Biochemistry, RAS, St. Petersburg, Russia
alex_shpakov@list.ru

Key words: allosteric regulation; G-protein coupled receptor; signal transduction; transmembrane domain; orthosteric agonist; luteinizing hormone receptor; pepducin; inverse agonist; receptor autoantibodies

αβγ-Heterotrimeric G-protein coupled receptors (GPCRs) play a key role in transmitting hormonal signals to intracellular signaling cascades, and are therefore responsible for the regulation of gene expression and post-translational modification of proteins, and are involved in the control of proliferation, apoptosis, and differentiation. Each GPCR has its own endogenous regulator, usually a hormone or neurotransmitter (orthosteric agonist), which binds with high affinity to the ligand-binding orthosteric site of the GPCR. This causes the receptor to transition to an active state and thereby triggers the signal transduction, mediated through different types of G-proteins and β-arrestins. Over the years, the main research in the field of GPCR pharmacology has been aimed at developing synthetic orthosteric site ligands with a specific spectrum of biological activity, as a result of which a significant part of the drugs that target GPCRs are ligands of their orthosteric site. Recently, the emphasis in the study of GPCRs has shifted towards the study of their allosteric regulation, and the main impetus for this was given by research in the field of deciphering the spatial structure of GPCRs, which started in the late 2000s, after the discovery of the 3D structure of the β2-adrenergic receptor [1]. At first, it was believed that the GPCR has relatively few allosteric sites, from one to 3–4. It was also assumed that the orthosteric and allosteric sites should be localized in different GPCR subdomains and not overlap. However, it was later shown that the number of allosteric sites in GPCRs is significantly greater; they are localized in all main subdomains of the receptor and can physically overlap with the orthosteric site [2].

The most typical localization of allosteric sites is the extracellular subdomains of GPCRs, including the outer vestibule of the transmembrane (TM) channel, formed by extracellular loops and the outer ends of hydrophobic TM helices [2, 3]. The functions of these subdomains are the primary recognition and low-affinity binding of hormones and the formation of homo- and heterodimeric receptor complexes [4]. In addition, some GPCRs, such as the receptors of luteinizing hormone (LH) and thyroid-stimulating hormone (TSH), have large ectodomains that contain high-affinity orthosteric sites. These ectodomains are targets for allosteric regulators, which not only modulate the binding characteristics of GPCRs, but also are capable of regulating their activity, as shown for autoantibodies to the TSH receptor [5, 6]. Endogenous regulators of extracellular allosteric sites, along with autoantibodies, can be receptor activity-modifying proteins (RAMPs), protomers of other receptors, including tyrosine kinase-type receptors and ionotropic receptors, as well as peptide regulators, for example, agouti-like peptide in the case of M3- and M4-melanocortin receptors. In recent years, low-molecular compounds have been developed that can bind to the cavity of the outer

vestibule of the TM channel [3]. They influence the accessibility of the orthosteric site for agonists and are able to regulate the activity of ligand-free GPCR.

Another localization of allosteric sites is the internal cavity of the TM channel, which is accessible to relatively small molecules and ions. Within this cavity there are usually several allosteric sites, which may overlap with each other and with the orthosteric site. Endogenous ligands of allosteric sites located in the TM channel are most often sodium ions and some divalent metals (Zn^{2+}, Mg^{2+}, Ca^{2+}, and Cu^{2+}) [3, 7]. Currently, many low-molecular weight compounds have been developed that are able to penetrate inside the TM channel and bind to allosteric sites located in it. Thus, to regulate the LH and TSH receptors, we have developed thieno[2,3-d]-pyrimidine derivatives with the activity of both full and inverse agonists and neutral antagonists of these receptors [3, 8, 9]. When administered to rats, they were active both by intraperitoneal and oral routes of delivery, indicating their stability and good absorption in the gastrointestinal tract. Low-molecular weight allosteric agonists did not compete with gonadotropins and TSH for binding to the receptor, thereby not interfering with the regulation of endocrine functions by endogenous hormones.

In the TM domain, allosteric sites can also be located on its lateral surfaces, in contact with the lipid phase of the membrane [2, 7]. Endogenous ligands of such sites can be plasma membrane lipids, primarily cholesterol and phospholipids, as well as some steroid hormones and their precursors (progesterone, pregnenolone, etc.). The interaction of lipids with these allosteric GPCR sites provides a close relationship between receptor activity, on the one hand, and the lipid composition and physicochemical properties of the membrane, on the other [3, 7].

Another localization of allosteric sites is the cytoplasmic loops and the TM channel vestibule oriented inward to the cell. The endogenous regulators of these sites are heterotrimeric G-proteins and β-arrestins, which not only function as hormonal signal transducers, but also carry out allosteric modulation of receptor activity, determining their affinity for the orthosteric agonist and the bias of intracellular signaling [3]. In recent years, there has been intensive development of synthetic allosteric regulators of intracellular sites – both low-molecular compounds and peptides modified by hydrophobic radicals, derivatives of functionally important intracellular sites of GPCR, called pepducins [10]. We have developed palmitate-modified pepducins NKDTKIAKK-Nle-A(562–572)-K(Palm)A and 612–627(Palm), corresponding to regions of the third cytoplasmic loop of the LH and TSH receptors, which *in vitro* and *in vivo* exhibited the activity of partial agonists of those receptors, derivatives of the primary structure of which they were [3]. The action of pepducins is based on a change in the conformation of intracellular allosteric sites due to their specific interaction with the complementary regions of the cytoplasmic loops of the receptor and the intracellular vestibule of the TM channel.

Along with the multiplicity of allosteric sites, it is necessary to take into account the wide range of pharmacological activities of their ligands. These ligands can increase or decrease the affinity of the orthosteric agonist for the receptor and the effectiveness of its action, being positive (PAM) or negative (NAM) allosteric modulators respectively [2, 7]. They can fine-tune signal transduction by biasing the action of the orthosteric agonist on intramolecular targets, and are then classified as silent allosteric modulators (SAMs). They can be endowed with their own activity as full/partial agonists, inverse agonists and neutral antagonists, or combine the properties of allosteric agonists and modulators (Ago-PAM, Ago-NAM).

Conclusion: Thus, the allosteric regulation of GPCRs is based on a multiplicity of allosteric sites that differ in localization, structural and functional organization, and endogenous regulators. These sites can influence not only the conformation and binding characteristics of the orthosteric site, but also interact with each other. Along with such endogenous ligands of allosteric GPCR sites as simple ions, lipids, amino acids, and polypeptides. The processes of formation of their complexes with G-proteins, β-arrestins, RAMPs, other adapter and regulatory proteins, as well as the formation of homo- and heterodi(oligo)meric receptor complexes are important in the allosteric regulation of GPCRs. All this significantly complicates and enriches the mechanisms of GPCR-mediated signal transduction, and creates a significant theoretical and practical basis for the creation of a large number of allosteric regulators that differ in their selectivity and pharmacological profile. It should be noted that such regulators, unlike orthosteric agonists, act more mildly, which reduces the risks of GPCR desensitization, and are also more specific to intracellular effectors, which avoids a number of side effects characteristic of hormonal therapy. For some types of GPCRs, for example, for the TSH receptor, for which there are no suitable pharmacological regulators of the orthosteric site, the creation of allosteric regulators is the only approach [3].

Funding: The study is supported by a grant from the Russian Science Foundation (project No. 19-75-20122).

Список литературы/References

1. Cherezov V. et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science*. 2007;318(5854):1258-1265. doi 10.1126/science.115057
2. Hedderich J.B. et al. The Pocketome of G-Protein-Coupled Receptors Reveals Previously Untargeted Allosteric Sites. *Nat Commun*. 2022;13:2567. doi 10.1038/s41467-022-29609-6
3. Shpakov A.O. Allosteric Regulation of G-Protein-Coupled Receptors: From Diversity of Molecular Mechanisms to Multiple Allosteric Sites and Their Ligands. *Int J Mol Sci*. 2023;24(7):6187. doi 10.3390/ijms24076187
4. Liu L. et al. Allosteric Ligands Control the Activation of a Class C GPCR Heterodimer by Acting at the Transmembrane Interface. *Elife*. 2021;10:e70188. doi 10.7554/eLife.70188
5. Duan J. et al. Structures of full-length glycoprotein hormone receptor signalling complexes. *Nature*. 2021;598(7882):688-692. doi 10.1038/s41586-021-03924-2
6. Kleinau G. et al. Structural-Functional Features of the Thyrotropin Receptor: A Class A G-Protein-Coupled Receptor at Work. *Front Endocrinol*. 2017;8:86. doi 10.3389/fendo.2017.00086
7. Grundmann M., Bender E., Schamberger J., Eitner F. Pharmacology of Free Fatty Acid Receptors and Their Allosteric Modulators. *Int J Mol Sci*. 2021;22:1763. doi 10.3390/ijms22041763
8. Derkach K.V. et al. Comparison of Steroidogenic and Ovulation-Inducing Effects of Orthosteric and Allosteric Agonists of Luteinizing Hormone/Chorionic Gonadotropin Receptor in Immature Female Rats. *Int J Mol Sci*. 2023;24(23):16618. doi 10.3390/ijms242316618
9. Derkach K.V. et al. The Study of Biological Activity of a New Thieno[2,3-D]-Pyrimidine-Based Neutral Antagonist of Thyrotropin Receptor. *Bull Exp Biol Med*. 2022;172(6):713-717. doi 10.1007/s10517-022-05462-x
10. Ortiz Zacarías N.V. et al. Intracellular Receptor Modulation: Novel Approach to Target GPCRs. *Trends Pharmacol Sci*. 2018;39:547-559. doi 10.1016/j.tips.2018.03.002

doi 10.18699/bgrs2024-3.2-16

Mechanistic insights into MARK4 inhibition by galantamine toward therapeutic targeting of Alzheimer's disease

Adnan M.[1], DasGupta D.[2], Anwar S.[3]*, Shamsi A.[4], Siddiqui A.J.[1], Snoussi M.[5], Bardakci F.[1], Patel M.[1], Hassan M.I.[3]

[1] *Department of Biology College of Science University of Hail, Hail, Saudi Arabia*

[2] *College of Pharmacy University of Michigan, Ann Arbor, MI, United States*

[3] *Centre for Interdisciplinary Research in Basic Sciences, New Delhi, India*

[4] *Centre of Medical and Bio-Allied Health Sciences Research Ajman University, Ajman, United Arab Emirates*

[5] *Research and Development Cell, Department of Biotechnology, Parul Institute of Applied Sciences, Parul University, Vadodara, India*

** email2saleha@gmail.com*

Key words: kinase inhibitors, acetylcholinesterase inhibitors, Alzheimer's disease, isothermal titration calorimetry, molecular dynamics simulation

Motivation and Aim: Hyperphosphorylation of tau is an important event in Alzheimer's disease (AD) pathogenesis, leading to the generation of "neurofibrillary tangles," a histopathological hallmark associated with the onset of AD and related tauopathies. Microtubule-affinity regulating kinase 4 (MARK4) is an evolutionarily conserved Ser-Thr (S/T) kinase that phosphorylates tau and microtubule-associated proteins, thus playing a critical role in AD pathology. The uncontrolled neuronal migration is attributed to overexpressed MARK4, leading to disruption in microtubule dynamics. Inhibiting MARK4 is an attractive strategy in AD therapeutics.

Methods and Algorithms: Molecular docking was performed to see the interactions between MARK4 and galantamine (GLT). Furthermore, 250 ns molecular dynamic studies were performed to investigate the stability and conformational dynamics of the MARK4–GLT complex. We performed fluorescence binding and isothermal titration calorimetry studies to measure the binding affinity between GLT and MARK4. Finally, an enzyme inhibition assay was performed to measure the MARK4 activity in the presence and absence of GLT.

Results: We showed that GLT, an acetylcholinesterase inhibitor, binds to the active site cavity of MARK4 with an appreciable binding affinity. Molecular dynamic simulation for 250 ns demonstrated the stability and conformational dynamics of the MARK4–GLT complex. Fluorescence binding and isothermal titration calorimetry studies suggest-ed a strong binding affinity. We further show that GLT inhibits the kinase activity of MARK4 significantly (IC50 = 5.87 μM).

Conclusion: These results suggest that GLT is a potential inhibitor of MARK4 and could be a promising therapeutic target for AD. GLT's inhibition of MARK4 provides newer in-sights into the mechanism of GLT's action, which is already used to improve cognition in AD patients.

Funding: This research was funded by the King Salman Center for Disability Research through Research Group No. KSRG-2022-058.

doi 10.18699/bgrs2024-3.2-17

Usnic acid derivative – Tdp1 inhibitor enhances the antitumor and antimetastatic effects of topotecan and normalizes hemopoiesis *in vivo*

Chernyshova I.A.[1], Kornienko T.E.[1]*, Chepanova A.A.[1], Zakharenko A.L.[1], Filimonov A.S.[2], Luzina O.A.[2], Dyrkheeva N.S.[1], Nikolin V.P.[3], Popova N.A.[3], Salakhutdinov N.F.[3], Lavrik O.I.[1]

[1] *Novosibirsk Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia*

[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *t.kornienko1995@gmail.com*

Key words: usnic acid derivatives; Tdp1 inhibitors; anticancer therapy; topotecan

Motivation and Aim: Tyrosyl-DNA phosphodiesterase 1 (Tdp1) is an important DNA repair enzyme and one of the causes of tumor resistance to topoisomerase 1 inhibitors such as topotecan [1–4]. Inhibitors of this Tdp1 in combination with topotecan may improve the effectiveness of therapy. In this work, we synthesized a derivative of usnic acid, which is a hybrid of its known derivatives: tumor sensitizers to topotecan [5]. New compound inhibits Tdp1 in the submicromolar concentration range and enhances the effect of topotecan on the metabolic activity of cells of various lines according to the MTT test, while being non-toxic. *In vivo* experiments, this compound not only sensitizes tumors of mice to the action of topotecan, but also normalizes the state of the peripheral blood of mice [6].

Methods and Algorithms: The inhibitory properties of the compound (AF-185) on the purified Tdp1 were studied using a technique developed previously by our team [7]. A 16-mer single-stranded oligonucleotide carrying a fluorophore at the 5′-end and a quencher at the 3′-end was used as a biosensor. When within the Förster radius, the quencher suppressed the fluorescence of the fluorophore. The quencher was removed by Tdp1, resulting in fluorescence emission.

The cytotoxic/antiproliferative properties of the synthesized compound were studied using a standard MTT test on cell lines A-549, HCT-116, HeLa, MCF-7, T98G, MRC-5 and HEK293A. The ability of the usnic acid derivative to enhance the effect of topotecan was studied on the same panel of cell lines using a standard MTT-test.

In *in vivo* experiments, we studied the ability of AF-185 to enhance the effect of topotecan on murine Lewis lung carcinoma (LLC) and Krebs-2 carcinoma. LLC model is the only reproducible syngeneic murine model for lung cancer. This model is anaplastic, highly tumorigenic, and immunologically compatible with the murine system [8–10]. Subcutaneous transplantation leads to the development of a primary node at the injection site and the appearance of metastases in the lungs on days 17–21 [11]. The effect of the drugs was evaluated at the end of the experiment by the weight and size of the tumor, the quantity of lung metastases.

Krebs-2 carcinoma has a high degree of malignancy. The ascitic form of Krebs-2 leads to death on 14–18 days. The tumor is weakly sensitive or insensitive to the action of cytostatics [12]. The effect of the treatment was evaluated at the end of the experiment

by the weight of the tumor, the concentration of tumor cells in the ascitic fluid. In addition, we calculated the concentration of erythrocytes and leukocytes in the blood of mice.

Results: Half maximal inhibitory concentration (IC_{50}) of the usnic acid derivative in relation to Tdp1 was 0.12 ± 0.01 μM. Whereas, semitoxic concentration (CC_{50}) of the compound in relation to cell lines (CC_{50}, μM) was > 100 μM. Usnic acid derivative turned out to be promising sensitizer of the action of topotecan on cultured cell lines (data is presented in Table 1).

Table 1. CC_{50} values for topotecan in the presence of 5 μM Tdp1 inhibitor (AF-185), μM

	MRC-5	HeLa	HEK293A	HCT-116	A-549	MCF-7	T98G
Tpc+185	>10	4.6 *	0.033	0.21 *	0.25 *	0.63 *	1.3

* Differences in the presence and absence of the Tdp1 inhibitor are significant according to the Mann–Whitney test at a minimum of two (or three) concentrations of topotecan, $p < 0.05$; ** nd – not determined. A 50 % inhibition of cell survival was not achieved.

In the Lewis lung carcinoma model, it was shown that the administration of topotecan (1 mg/kg intraperitoneally, i/p) reduced primary node size; the tumor growth inhibition (TGI) value was 36.4 %, compared to intact control. The use of the combination of topotecan with 50 mg/kg AF-185 (intragastrical administration of AF-185, i/g) leads to a further reduction in tumor weight, TGI was 47.7 %. The administration of AF-185 i/g by itself did not lead to a decrease in tumor weight. The best results were obtained for the combination of topotecan with AF-185 (50 mg/kg, i/p), TGI was 57.2 %; the difference was significant both in comparison with both controls and with the topotecan group ($p < 0.05$). The effect of AF-185 on the antimetastatic properties of topotecan was less pronounced. At the dose used, topotecan had virtually no effect on the number of metastases in the lungs, and the number of metastases was not affected by the use of AF-185 both i/p and i/g, as well as the combination of topotecan with AF-185 i/g. The combination of topotecan with AF-185 i/p reduced the number of metastases, although the difference with the control and with the topotecan group was insignificant. Thus, the use of AF-185 in mono mode has no effect on either the growth of the primary node or the number of metastases in the lungs. In combination with topotecan, there was a tendency to reduce the weight of the primary tumor and the number of metastases, which was more pronounced with intraperitoneal administration of AF-185.

In the Krebs-2 ascitic carcinoma model, it was shown that the combination of topotecan and AF-185 i/p was more effective in terms of ascites weight and the number of tumor cells in ascitic fluid than the same combination with AF-185 i/g. Also, the combination of topotecan and AF-185 i/p had a more pronounced antitumor effect than topotecan, AF-185 i/p, and AF-185 i/g separately. Mice with Krebs-2 carcinoma (intact control) have an increased number of leukocytes and a decreased number of erythrocytes compared to healthy mice. An increase in white blood cell count in the peripheral blood is associated with the inflammatory process that develops during tumor formation and growth. The administration of solvent for AF-185 (DMSO + Tween-80), topotecan alone, as well as topotecan in combination with AF-185 i/g, slightly reduced the number of leukocytes. It should be noted that topotecan at the dose of 1 mg/kg was not hemotoxic and partly normalized the concentrations of leukocytes and erythrocytes, apparently due to its antitumor effect. Treatment of mice with only AF-185, regardless of the route of administration, reduced the number of leukocytes somewhat more effectively than the drugs listed above, but the decrease was significant only in comparison with the control

without treatment. Administration of topotecan in combination with AF-185 i/p significantly reduced the number of leukocytes compared to the control without treatment, the control with DMSO +Tween-80 and the group of mice receiving only topotecan.

Mice with Krebs-2 carcinoma have a reduced number of red blood cells compared to healthy mice, which may be caused by general intoxication of the body. The number of red blood cells practically did not change with the introduction of the solvent. At the same time, treatment with drugs and their combinations significantly increased the number of erythrocytes, regardless of the method of administration of AF-185. The largest effect was again exerted by the combination of topotecan and AF-185 i/p, which increased the number of erythrocytes to healthy control. Thus, the blood cell counts returned to normal values (red blood) or significantly improved (white blood) after treatment with combination Tpc + AF-185 i/p, i.e., hematopoiesis was normalized.

Conclusion: Thus, AF-185 has antitumor and antimetastatic effects, normalizes hemopoiesis and may be the prototype of a new class of additional therapy for cancer.

Funding: The study is supported by the Russian Scientific Foundation grant No. 23-74-01078.

References

1. Comeaux E.Q., van Waardenburg R.C. Tyrosyl-DNA phosphodiesterase I resolves both naturally and chemically induced DNA adducts and its potential as a therapeutic target. *Drug Metab Rev*. 2014;46:494-507. doi 10.3109/03602532.2014.971957
2. Brettrager E.J. et al. Tyrosyl-DNA Phosphodiesterase I N-Terminal Domain Modifications and Interactions Regulate Cellular Function. *Genes*. 2019;10:897. doi 10.3390/genes10110897
3. Alagoz M. et al. DNA repair and resistance to topoisomerase I inhibitors: Mechanisms, biomarkers and therapeutic targets. *Curr Med Chem*. 2012;19:3874-3885. doi 10.2174/092986712802002590
4. Pommier Y. et al. Tyrosyl-DNA-phosphodiesterases (TDP1 and TDP2). *DNA Repair*. 2014;19:114-129. doi 10.1016/j.dnarep.2014.03.020
5. Zakharenko A.L. et al. Usnic Acid Derivatives Are Effective Inhibitors of Tyrosyl-DNA Phosphodiesterase 1. *Russ J Bioorg Chem*. 2017;43(1):84-90. doi 10.1134/S1068162017010125
6. Kornienko T.E. et al. Enhancement of the Antitumor and Antimetastatic Effect of Topotecan and Normalization of Blood Counts in Mice with Lewis Carcinoma by Tdp1 Inhibitors – New Usnic Acid Derivatives. *Int J Mol Sci.* 2024;25(2):1210. doi 10.3390/ijms25021210
7. Zakharenko A.L. et al. Natural Products and Their Derivatives as Inhibitors of the DNA Repair Enzyme Tyrosyl-DNA Phosphodiesterase 1. *Int J Mol Sci.* 2023;24:5781. doi 10.3390/ijms24065781
8. Bertram J.S., Janik P. Establishment of a cloned line of Lewis lung carcinoma cells adapted to cell culture. *Cancer Lett*. 1980;11:63-73
9. Zhu H. et al. A simple bioluminescence imaging method for studying cancer cell growth and metastasis after subcutaneous injection of Lewis lung carcinoma cells in syngeneic C57BL/6 mice. *React Oxyg Species*. 2018;5:118-125. doi 10.20455/ros.2018.813
10. Berdel W.E. Ether lipids and analogs in experimental cancer therapy. A brief review of the Munich experience. *Lipids.* 1987;22:970-973
11. Ma X.M. et al. Comparison of mouse models of Lewis lung carcinoma subcutaneously transplanted at different sites. *Acta Lab Anim Sci Sin*. 2017;25:386-390
12. Поттер Е.А. и др. Характеристика режимов терапевтического воздействия циклофосфана и препаратов двуцепочечной ДНК на опухоль Кребс-2, растущую в асцитной форме, приводящих к эрадикации первичного асцита. *Вавиловский журнал генетики и селекции*. 2016;20(1):108-124. doi 10.18699/VJ15.117

doi 10.18699/bgrs2024-3.2-18

In vitro cytotoxic potential in human cancer cell line (C33a HPV-negative counterparts cervical cancer cell) of plant alkaloid, terpenoid and steroid derivatives

Hamad M.S.[1], Pokrovskii M.A.[1], Hamad S.S., Usenov K.[1], Pokrovskii A.G.[1]., Shinkarenko E.M.[1, 2], Finke A.O.[1, 2], Mironov M.E.[1, 2], Shults E.E.[2]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Novosibirsk Institute of Organic Chemistry, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia*

** mohammed_s1983@yahoo.com*

Key words: Medicinal plant; Anticancer; Antitumor; Cytotoxicity; Herbal medicine; apoptosis

Motivation and Aim: A tumor is a pathological process based on unregulated cell reproduction [1]. Cervical cancer is a malignant tumor of the mucous membrane of the vaginal portion and/or cervical canal. In 2023, malignant neoplasms claimed the lives of around 10 million individuals, with prevalent cancers encompassing breast, lung, colon and rectum, non-melanoma skin tumors, prostate, stomach, and cervical cancers. Cervical cancer, with approximately 16.5 thousand new cases annually in Russia, ranks fifth among cancers affecting women. Detection rates in early stages stand at 65% of patients, predominantly impacting middle-aged women. The historical use of herbal products in medicine, including anticancer drugs, remains significant [2]. Clinical trials have demonstrated the efficacy of various natural compounds like alkaloids, terpenes, phenols, and flavonoids [3]. Plant-derived substances form a substantial class of antitumor drugs, intervening in metabolic processes and enzymes crucial for their effects. Plant alkaloids, terpenoids, steroids and their synthetic derivatives, disrupt DNA integrity, transcription, repair, and mitosis, halting tumor cell proliferation [4, 5]. This report provided some data on the antitumor effects and mechanisms of alkaloids, terpenoids and steroids in the context of cervical cancer's hallmarks. A comparative analysis efficacy in cervical tumors cell line (C33a) in vitro. The paper provides an overview of the anticancer properties In vitro cytotoxic potential in human cell line (C33a HPV-negative counterparts), cervical cancer cell effects of derived plant alkaloids, terpenoids and steroids, emphasizing their pivotal role in the therapeutic landscape against cancer in future. The aim is to study the cytotoxic activity and the molecular mechanism of apoptosis-inducing effects of plant terpenoids, alkaloids and steroids in human tumor cells.

Materials: Human cell line cervical cancer (C33a HPV-negative counterparts), 148 compounds of derivatives of alkaloids, terpenoids and steroids, obtained in the laboratory of medical chemistry (No. 13-LMX). (Head of the laboratory, doctor of Chemical Sciences, Professor Shults Elvira Eduardovna). A mother (stock) solution of compounds at a concentration of 10 mM was obtained by dissolution in dimethyl sulfoxide and subsequently stored at -20 °C. FA, CHK, C -are alkaloids, LAN are derivatives of anthranilic acid (aromatic compounds), SV, Path - are sesquiterpenoids, HC – anthraquinone derivatives, KHAR, PD, PA, PA, – diterpenoids and MM –

Spirostanes, diosgenin derivatives (derivatives of steroid compounds). That is, all groups of substances are obtained by modification of natural compounds.
Method: *Cytotoxicity analysis of alkaloids, terpenoids and steroids using MTT test*
The cytotoxicity of alkaloids, terpenoids and steroids was evaluated using MTT test on potential in human cell line (C33a HPV-negative counterparts), cervical cancer carcinoma). According to stander protocol. Analysis of apoptosis by flow cytometry According to manufacturer's protocol.
Results: Study of the cytotoxic activity of plant derivatives, alkaloids, terpenes and steroids testing for cytotoxic activity in tumor cells humans was carried out in cell cultures C33a. Quantitative characteristics of cytotoxic activity, GI50, GI80 and GI90, as a result, it was discovered that some compounds are more active against 50% and 90% growth inhibition in cell lines C33a HPV-negative. C33a HPV-negative (MM-763, MM-765, MM-770, MM-771, MM-774, MM-775, MM-776, MM-784, MM-786, MM-787, MM-788, MM-789**,** MM-811, SHE-011**,** SHE052, SHE-022, SHE-027, SHE062, SHE078, FA-DM-13-1, FA-6921, FA-DM-648, FA-685, CK-24-11, CK-24-46, CK-28-84, MM-8092). Studying the ability of the most active compounds of terpenes, alkaloids, and steroids to induce apoptosis in C33a HPV-negative cells line in vitro.The cytotoxicity assay results obtained in the first phase of this project show high activity of a number of alkaloids, terpenes and steroids in suppressing cancer cell growth. We chose the most effective method for studying the type of cell death induction - cytometry, to search for the most active plant compounds alkaloids and steroids (**SHE052, FA-DM-13-1, CK-24-46, MM-763, MM-771, MM-811**) in two concentration **GI_{50}, µM**, **GI_{90}, µM.** Labeling cells with annexin-FITC markers makes it possible to detect early apoptosis (expression of phosphatidylserine in outer leaf of the cytoplasmic membrane and binding of annexin to phosphatidylserine) and propidium iodide – late cell death (fragmentation of nuclear DNA and penetration of propidium iodide through holes in the cytoplasmic and nuclear membranes and binding to the nuclear DNA).
Annexin V binds to phosphatidylserine. Usually, phosphatidylserine is found only in the inner layer of the cell membrane, but at the very beginning of apoptosis, PtdSer (phosphatidylserine) is transferred to the outer layer leaf. There it can bind the annexin V protein in a Ca2+-dependent manner. Thus, the early apoptosis phase is easily detected by cell labeling annexin V-FITC conjugate using flow cytometry (FACS).
Discussion: It has been reported that natural products steroids, terpenoids and alkaloids from plant sources have different biological activity such as antioxidant, anti-inflammatory or antiproliferative. Many of the reported anticancer effects natural compounds also target cellular proteins that play important role in signal transduction, apoptosis or cell cycle arrest. Our study reports on exploring the potential cytotoxic/antiproliferative activity of 148 secondary compounds metabolites found in plants such as alkaloids, steroids and terpenes, in the fight against tumor cells C33a HPV-negative. It has been shown that some of these plant-derived compounds have different mechanisms for the destruction of cancer cells, such as cell arrest cycle and intensification of the process of apoptotic cell death. These discoveries may pave the way for the development of new cancer treatments that exploit the therapeutic potential of natural compounds. Was found that C33a HPV-negative (Cervical cancer) cells exposed to effects of MM-763 and FA-DM-13-1, steroid derivatives, show high decreased metabolic activity (MTT test), externalization phosphatidylserine (Annexin-V). Of the 148 compounds studied from among derivatives of alkaloids, terpenoids and steroids, as shown in this work, 27 compounds showed high cytotoxic activity against

one or several cancer cell lines (C33a HPV-negative, 2 of these 27 compounds showed induction of apoptosis in a cell line (C33a HPV-negative counterparts), cervical cancer cell. Cells treated with compounds for (6) hours **GI_{50}, µM** MM-763,/ FA-DM-13-1 (at concentrations of 2.9 µM, 3.5 µM) contained 2.2%, 0.1%, 10.5% / 0.1%,0.1%, 0.5% cells in the stage of early apoptosis, respectively and **GI90, µM** MM-763, FA-DM-13-1 (at concentrations of 13 µM, 80 µM). contained 1.5%, 1.5%, 10.7% / 55.0% cells in the stage of early apoptosis, respectively, for 12 hours **GI_{50}, µM** MM-763, FA-DM-13-1 (at concentrations of 2.9 µM, 3.5 µM) contained 0.3%, 0.4%, 22.0% / 0.4%, 0.4% , 1.4% cells in the stage of early apoptosis, respectively and **GI90, µM** MM-763, FA-DM-13-1 (at concentrations of 13 µM, 80 µM), contained 2.0%,4.1%,67.1% / 0.6%, 6.7%, 78.5% cells in the stage of early apoptosis, respectively, for 24 hours **GI_{50}, µM** MM-763, FA-DM-13-1 (at concentrations of 2.9 µM, 3.5 µM), contained 0.4%, 1.0%,70.0% / 0.0%, 0.0% , 0.7% cells in the stage of early apoptosis, respectively and **GI90, µM** MM-763, FA-DM-13-1 (at concentrations of 13 µM, 80 µM), contained 0.8%, 0.8%, 69.2% / 0.0%, 0.2%, 48.3% cells in the stage of early apoptosis respectively.

References

1. American Cancer Society. Cancer Facts & Figures 2016. Atlanta, GA, USA: American Cancer Society, 2016
2. Ye H., Wang L., Ma L., Ionov M., Qiao G., Huang J., Cheng L., Zhang Y., Yang X., Cao S., Lin X. Protein kinases as therapeutic targets to develop anticancer drugs with natural alkaloids. *Front Biosci (Landmark Ed).* 2021;26(11):1349-1361
3. Hasanpourghadi M. et al. Phytometabolites targeting the warburg effect in cancer cells: A mechanistic review. *Curr Drug Targets*. 2017;18(9):1086-1094
4. Dey P., Kundu A., Chakraborty H.J., Kar B., Choi W.S., Lee B.M., Bhakta T., Atanasov A.G., Kim H.S. Therapeutic value of steroidal alkaloids in cancer: current trends and future perspectives. *Int J Cancer*. 2019;145(7):1731-1744
5. Huang Y., Li G., Hong C., Zheng X., Yu H., Zhang Y. Potential of steroidal alkaloids in cancer: Perspective insight into structure-activity relationships. *Front Oncol*. 2021;11:733369

doi 10.18699/bgrs2024-3.2-19

Metabolic response promotes systematically the anti-inflammatory effect of ginsenosides

He M.[1]*, Sun Q.[1], Dou J.[1], Yu M.[1], Li Y.[2], Sun M.[1]**

[1] *Changchun University of Chinese Medicine, Changchun, China*

[2] *Changchun Sino-Russian Science and Technology Park Co., Ltd., Changchun, China*

** heminccucm@hotmail.com; **sunm2000@hotmail.com*

Key words: Ginseng ginsenosides; anti-inflammation; metabolic regulations; zebrafish

Motivation and Aim: The use of bioactive compounds to induce metabolic reprogramming is emerging as a novel adjuvant strategy for clinical immunotherapy. Gingseng ginsenosides, exists in ginseng roots, have shown anti-inflammatory activity by functioning as a glucocorticoid receptor agonist, but without glucocorticoid-like side effect to inhibit wound healing. Considering the systematic influence of glucocorticoid to the body, understanding the relationship between the anti-inflammatory functions and metabolic effects of Gingseng ginsenosides, which may differ from glucocorticoids, may help understand more biological mechanism details of Gingseng ginsenosides, as well as identify metabolic biomarkers associated with anti-inflammation. Therefore, we aim to investigate the endogenous metabolic response of Gingseng ginsenosides underlying the anti-inflammatory response in zebrafish model, providing a rationale for incorporating Ginseng into immune-modulating nutraceuticals.

Methods and Algorithms: The anti-inflammatory effect of Gingseng ginsenosides was examined by analyzing neutrophil cell counts under a fluorescent microscope and measuring inflammatory cytokine gene expression using q-PCR techniques in zebrafish larvae under inflammatory induction. The metabolic impact of Gingseng ginsenosides, particularly on primary metabolites, was assessed using high performance liquid chromatography coupled with mass spectrometry, and metabolic pathways were analyzed using the KEGG pathway database.

Results: Amputation induced the infiltration of neutrophils and macrophages towards the amputated edges, accompanied by upregulation of gene expression associated with pro-inflammatory cytokines. Gingseng ginsenosides were equally effective in alleviating inflammatory responses in injured zebrafish as beclomethasone, but with different metabolic responses, particularly in fatty acid metabolism and downstream aromatic amino acids in the TCA cycle.

Conclusion: Gingseng ginsenosides shows promise as a drug candidate for treating inflammatory responses and as a valuable supplement for enhancing immune regulation.

Funding: This research was supported by the National Natural Science Foundation of China (No. 82004030); the Scientific and Technological Developing Project of Jilin Province (No. 20210402042GH); the Jilin Provincial Development and Reform Commission (No. 2023C028-1); the Scientific and Technological Developing Project of Jilin Province (No. YDZJ202101ZYTS119); the Scientific and Technological Developing Project of Jilin Province (No. YDZJ202301ZYTS151).

doi 10.18699/bgrs2024-3.2-20

Structure-guided identification of Baicalin exhibiting anti-Alzheimer's effects via CLK1 inhibition

Khan S., Hassan M.I., Islam A.*

Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India
** aislam@jmi.ac.in*

Key words: Alzheimer's disease; kinase inhibition; CDC2-like kinase 1; hyperphosphorylation; structure-guided drug discovery

Motivation and Aim: Alzheimer's disease (AD) is a rapidly increasing neurodegenerative disorder, projected to affect 78 million people by 2050 [1]. Current treatments provide only modest symptomatic relief [2]. Tau protein hyperphosphorylation is a key pathological feature of AD, and evidence suggests that targeting tau phosphorylation through the inhibition of protein kinases, particularly CLK1 (CDC2-like kinase 1), could reduce tau aggregation and neuronal death [3]. This study aims to investigate the potential of Baicalin as a CLK1 inhibitor, offering a novel therapeutic approach for AD and other tauopathies.

Methods and Algorithms: We employed structure-based molecular docking to predict the binding affinity of Baicalin to CLK1. Molecular dynamics simulations were conducted to analyze the stability and interactions of the Baicalin-CLK1 complex. Fluorescence spectroscopy was used to confirm the binding mechanism. Additionally, a kinase inhibition assay was performed to determine the inhibitory potential of Baicalin on CLK1, quantified by IC50 values.

Results: Our molecular docking studies revealed a strong binding affinity of Baicalin to CLK1. Molecular dynamics simulations confirmed the stability of the Baicalin-CLK1 complex. Fluorescence spectroscopy validated the binding interactions between Baicalin and CLK1. The kinase inhibition assay demonstrated that Baicalin significantly inhibits CLK1 activity, with an IC50 value of 12.04 μM.

Conclusion: Baicalin exhibits significant potential as a CLK1 inhibitor, effectively reducing CLK1 activity in vitro. These findings position Baicalin as a promising lead molecule for developing specific CLK1 inhibitors, offering a novel therapeutic strategy for AD and other tauopathies. Further research and clinical trials are necessary to validate these findings and explore the therapeutic efficacy of Baicalin *in vivo*.

Funding: Sumaiya Khan acknowledges the DST INSPIRE Fellowship with reference number: IF220384.

References

1. Li X. et al. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2019. *Front Aging Neurosci*. 2022;14:937486. doi 10.3389/fnagi.2022.937486
2. Khan S. et al. Nature's toolbox against tau aggregation: An updated review of current research. *Ageing Res Rev*. 2023;87:101924. doi 10.1016/j.arr.2023.101924
3. Jain P. et al. Human CDC2-like kinase 1 (CLK1): A novel target for Alzheimer's disease. *Curr Drug Targets*. 2014;15(5):539-550. doi 10.2174/1389450115666140226112321

doi 10.18699/bgrs2024-3.2-21

Systematic regulation of essential oil extracted from *Perilla Frutescens L.* against inflammation in zebrafish model: combining the transcriptomic and metabolomic analysis

Mengmeng Sun[1]*, Yao Fu[1], Jie Cheng[1], Lulu Wang[1], Yongping Li[2], Min He[1]**

[1] *Changchun University of Chinese Medicine, Changchun, China*

[2] *Changchun Sino-Russian Science and Technology Park Co., Ltd., Changchun, China*

** sunm2000@hotmail.com; ** heminccucm@hotmail.com*

Key words: essential oil; anti-inflammation; metabolic regulations; zebrafish

Motivation and Aim: Belonging to the Lamiaceae family, *Perilla frutescens L.* is extensively cultivated across Asian countries including China. Known as s a "drug homologous food" with aromatic taste, its leaves have been widely used as a culinary herb and a representative flavor and food agent, mainly owing to a compelling aromatic profile in its essential oils. However, the chemical components in relation to the pharmacological potentials of its essential oils (PFO) has not been well understood currently. Therefore, our aim of this study is to investigate the chemical composition and potential anti-inflammation effect of PFO and to understand the body response of its function by combining the transcriptomics and metabolomics data, emplying in a tail-fin amputated zebrafish model.

Results: The PFO contains high concentration of Perilla ketone (42.41 %), significantly inhibited neutrophils migration to the amputation site, and inhibited the gene expressions of pro-inflammatory cytokines TNF-α, IL-6 and IL-1β. The combined analysis of transcriptomics and metabolomics revealed that the regulatory inflammation of the PFO is closely related to amino acid metabolism and oxylipins pathways represented by prostagladins.

Conclusion: The PFO shows potentials as anti-inflammatory agent as a valuable supplement, which provides an important reference for its development value and rational utilization as functional food in the future.

Funding: This research was supported by the National Natural Science Foundation of China (No. 82004030); the Scientific and Technological Developing Project of Jilin Province (No. 20210402042GH); the Jilin Provincial Development and Reform Commission (No. 2023C028-1); the Scientific and Technological Developing Project of Jilin Province (No. YDZJ202101ZYTS119); the Scientific and Technological Developing Project of Jilin Province (No. YDZJ202301ZYTS151).

doi 10.18699/bgrs2024-3.2-22

Sequence-based computational protein analysis MDR pumps for use in epidemiology and drug design

Nazarov P.A.
Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia
mvk752002@gmail.com

Key words: alignment; MDR pump; AcrAB-TolC; sequence, epidemiology

Motivation and Aim: Antibiotic resistance is a pressing problem in modern healthcare. Attempts to predict the effect of a drug on specific strains usually do not lead to positive results. We chose to model MDR pump genes known for their variability. Multidrug-resistant (MDR) pumps are essential for protecting bacteria from antibiotics. These pumps offer a basis for the non-specific defense of bacteria and are categorized into six families of MDR pumps [1].
Methods and Algorithms: To solve our problem of predicting the right antibiotic for treatment, we use methods from molecular microbiology, sequence alignment, molecular docking and molecular biology, which allows us to obtain reliable results.
Results: Using molecular microbiology, we validate data obtained through sequence analysis. Using molecular biology methods, strains will be created to create pathogen resistance systems in the laboratory. Using molecular docking, we study how the mutation process can affect the antibacterial substances we are developing.
Conclusion: We have developed a pipeline for analyzing pump sequences. This allows us to predict resistance to the antibiotics we are developing based on the primary structure of a set of certain bacterial pump genes.
Funding: This work was supported by grant from the Russian Science Foundation 22-15-00 099.

References

1. Nazarov P.A. MDR Pumps as Crossroads of Resistance: Antibiotics and Bacterio-phages. *Antibiotics* (*Basel*). 2022;11(6):734. doi 10.3390/antibiotics11060734

doi 10.18699/bgrs2024-3.2-23

Computational screening of repurposed drugs for HMG-CoA synthase 2 in Alzheimer's disease

Shamsi A.[1]*, Furkan M.[2], Khan M.S.[3], Yadav D.K.[4], Shahwan M.[1]

[1] *Center for Medical and BioAllied Health Sciences Research Ajman University UAE*

[2] *Department of Biochemistry Aligarh Muslim University, Aligarh, India*

[3] *Department of Biochemistry College of Science King Saud University KSA*

[4] *Gachon Institute of Pharmaceutical Science and Department of Pharmacy College of Pharmacy, Gachon University Incheon, Incheon, Republic of Korea*

** m.shamsi@ajman.ac.ae*

Key words: Human mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase 2; Alzheimer's disease; drug repurposing; small molecule inhibitors; virtual screening

Motivation and Aim: HMGCS2 (mitochondrial 3-hydroxy-3-methylglutaryl-COA synthase 2) plays a pivotal role as a control enzyme in ketogenesis, and its association with the β-amyloid precursor protein (APP) in mitochondria implicates a potential involvement in Alzheimer's disease (AD) pathophysiology. Our study aimed at identifying repurposed drugs using the DrugBank database capable of inhibiting HMGCS2 activity.

Methods and Algorithms: Exploiting the power of drug repurposing in conjunction with virtual screening and molecular dynamic (MD) simulations against predefined targets, we present new in-silico insight into structure-based drug repurposing.

Results: The initial molecules were screened for their binding affinity to HMGCS2. Subsequent interaction analyses and extensive 300 ns MD simulations were conducted to explore the conformational dynamics and stability of HMGCS2 in complex with the screened molecules, particularly Penfluridol and Lurasidone.

Conclusion: The study revealed that HMGCS2 forms stable protein-ligand complexes with Penfluridol and Lurasidone. Our findings indicate that Penfluridol and Lurasidone competitively bind to HMGCS2 and warrant their further exploration as potential repurposed molecules for anti-Alzheimer's drug development.

Funding: The study is supported by Ajman University.

doi 10.18699/bgrs2024-3.2-24

Adrenochrome formation during photochemical transformation of tailored epinephrine derivatives

Starodubtseva E.[1], Karogodina T.[1, 2], Panfilov M.[1, 2], Vorob'ev A.[1, 2], Moskalensky A.[1]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia*

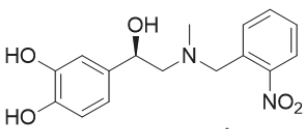
Key words: adrenochrome; epinephrine; photoremovable protecting group

Motivation and Aim: Adrenergic receptors are a class of G protein-coupled receptors that mediate the physiological effects of catecholamines adrenaline (epinephrine) and noradrenaline (norepinephrine). These receptors are widely distributed throughout the body, and their activation triggers various responses, including changes in heart rate, smooth muscle contraction, and metabolic processes.

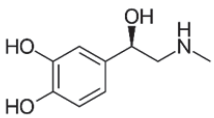
30 years ago, Muralidharan and Nerbonne reported a family of photolabile compounds capable of the release of adrenergic receptor agonists under UV light [1]. This ability is useful for light-induced modulation of receptor activity [2–4]. The compounds were prepared using 2-nitrobenzyl or substituted 2-nitrobenzyl photoremovable protecting groups. It was shown that photorelease is faster when the protective group is attached to the β-amino group rather than one of the phenolic oxygens of the catecholamine.

However, the detailed studies of the photolysis products are lacking, especially for the epinephrine analogs. On the other hand, the photooxidation of epinephrine is a common effect, and we aimed at the detailed study of the photolysis of its photolabile analogs.

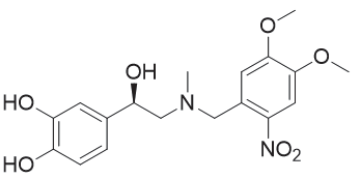
Methods and Algorithms: We have prepared two photolabile analogs of epinephrine, one previously described (compound 1) and one novel but structurally similar to the reported agonists (compound 2). The additional methoxy groups on the latter are aimed at shifting the excitation spectra to longer wavelengths and increasing the photorelease quantum yield. The compounds were illuminated by UV LEDS, and the photolysis products were analysed by UV-Vis spectroscopy and by high-performance liquid chromatography (HPLC).



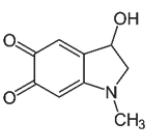
Compound 1



Epinephrine



Compound 2



Adrenochrome

Results: During the photolysis, we observed rapid and profound change of the absorption spectra of the compounds, which should be accompanied by the liberation of free epinephrine. Surprisingly, novel peak certainly appears in longer-wavelength region around 500 nm. HPLC analysis revealed that this product has absorption peaks at 220, 300 and 485 nm, which is characteristic for adrenochrome.

Adrenochrome is a chemical compound produced by the oxidation of epinephrine. It was the subject of limited research from the 1950s through to the 1970s as a potential cause of schizophrenia [5]. Some studies reported hallucinogenic and psychoactive properties of the compound. It has also been implicated in cardiotoxicity [6] and generally is a cytotoxic molecule. Adrenochrome can be produced by photooxidation of epinephrine. However, the process is slow and requires hours of irradiation. In contrast, the speed of adrenochrome generation observed in our experiments is quite high and cannot be explained by the photooxidation alone.

Conclusion: Our results indicate that aminochromes may be formed during photorelease of catecholamines from their photolabile analogs, and it should be tested before biological applications.

Funding: The study was supported by the Russian Science Foundations (grant #23-75-10049).

References

1. Muralidharan S., Nerbonne J.M. Photolabile “caged” adrenergic receptor agonists and related model compounds. *J Photochem Photobiol, B Biol.* 1995;27:123-137. doi 10.1016/1011-1344(94)07063-T
2. Ellis-Davies G.C.R. Caged compounds: photorelease technology for control of cellular chemistry and physiology. *Nat Methods.* 2007;4:619-628. doi 10.1038/nmeth1072
3. Vorobev A.Yu., Moskalensky A.E. Long-wavelength photoremovable protecting groups: On the way to in vivo application. *Comput Struct Biotechnol J*. 2020;18:27-34. doi 10.1016/j.csbj.2019.11.007
4. Litunenko D.N., Karogodina T.Y., Vorob’ev A.Y., Moskalensky A.E. Novel UV-Releaser of Arachidonic Acid Based on Nitrophenyl-type Photoremovable Protecting Group. *J Biomed Photonics Eng*. 2022;8:030305. doi 10.18287/JBPE22.08.030305
5. Smythies J. The adrenochrome hypothesis of schizophrenia revisited. *Neurotox Res*. 2002;4:147-150. doi 10.1080/10298420290015827
6. Behonick G.S., Novak M.J., Nealley E.W., Baskin S.I. Toxicology update: the cardiotoxicity of the oxidative stress metabolites of catecholamines (aminochromes). *J Appl Toxicol*. 2001;21:S15-S22. doi 10.1002/jat.793

doi 10.18699/bgrs2024-3.2-25

An analysis of virus-host interaction: case study for HIV-infection

Tarasova O.*, Biziukova N., Rudik A., Ivanov S., Poroikov V.
Institute of Biomedical Chemistry, Moscow, Russia
* *olga.a.tarasova@gmail.com*

Key words: virus-host interactions; HIV; bioinformatics; systems biology; data analysis; knowledgebase

Motivation and Aim: Analysis of virus-host interactions is essential for deep understanding of the molecular interplay that is caused by a viral infection. The large amount of experimental data available on people with viral infections can be used to develop computational models of virus-host interactions [1]. Such models can help to determine the mechanisms of viral infection progression and the efficacy of antiviral therapy, in order to improve existing and develop new treatment strategies including those that take into account the individual characteristics of a patient. Systems biology and bioinformatics approaches allow us to analyze the entire set of experimental data in silico and generate new knowledge.
Methods and Algorithms: The developed approach includes a set of methods for automated selection of relevant data from different sources and their integration. Based on the retrieved data, we performed a search for key regulators of HIV infection progression and carried out their experimental validation. We have developed algorithms to estimate the effectiveness of antiretroviral therapy, taking into account i.e. amino acid and nucleotide sequences of the virus. In particular, we used the in-house developed algorithms of named entity recognition and extraction of associations between extracted entities [2, 3] to perform the text and data mining of virus-host interactions. The developed algorithms of machine-learning based HIV drug resistance prediction [4, 5] was used for building web-based models aimed at drug resistance prediction [5]. Transcriptome-based bioinformatics analysis of peripheral blood mononuclear cells (PBMC) obtained from HIV-infected people (PLWH) with various viral infection progression and viral infection control allowed us to identify the molecular mechanisms of HIV infection progression that are influenced by individual characteristics of patient's reply on the infection [1]. An analysis of relationship between differential expression in genes of patients with HIV-infection and the therapeutic outcome was compared with experimental data resulting from clinical study on eleven people observed before starting antiretroviral therapy and 24 weeks after.
Results: The developed methods are integrated into the hiv-host platform (https://www.way2drug.com/hiv-host), which includes (1) databases on the interaction of HIV macromolecules with the human body and information on the history of HIV therapy, and (2) models and web applications for estimating viral drug resistance and efficacy of antiretroviral therapy, synergy of antiretroviral drugs, models for identifying master regulators of infection progression.
As a result of gene enrichement analysis performed for the samples obtained from PLWH, it was found that the elevated expression of proapoptotic and innate immunity genes before taking antiretroviral therapy can be associated with higher effectiveness of antiretroviral therapy. Some of observations found as a result of text and data mining

analysis [6] and bioinformatics evaluation of gene expression in PBMC of PLWH were confirmed in the clinical study.

Based on the developed methodology we expanded the data analysis on a set of viruses that included Hepatitis B and C viruses, SARS-CoV-2, influenza A and B, and herpes simplex virus. We collected data on interaction for three interactions on these viruses with (1) the host (human body), (2) with potential antivirals and (3) on the interaction of the host proteins with chemical compounds. Based on the performed data analysis we created freely available knowledgebase on the interaction of chemical compounds with viral proteins and their host targets.

Conclusions: The developed platform integrates information about the interaction of viruses and the host at the molecular level and provides the opportunity for further bioinformatic searches in order to obtain new knowledge about the pathogenesis of viral infections and the rational design of antiviral drugs.

Funding: The study is supported by the Program for Basic Research in the Russian Federation for a long-term period 2021–2030 (project No. 122030100170-5).

References

1. Ivanov S.M., Tarasova O.A., Poroikov V.V. Transcriptome-based analysis of human peripheral blood reveals regulators of immune response in different viral infections. *Front Immunol.* 2023;14:1199482. doi 10.3389/fimmu.2023.1199482
2. Tarasova O.A., Rudik A.V., Biziukova N.Y. et al. Chemical named entity recognition in the texts of scientific publications using the naïve Bayes classifier approach. *J Cheminform.* 2022;14:55. doi 10.1186/s13321-022-00633-4
3. Biziukova N. et al. Automated Extraction of Information From Texts of Scientific Publications: Insights Into HIV Treatment Strategies. *Front Genet.* 2020;11:618862. doi 10.3389/fgene.2020.618862
4. Tarasova O. et al. A Computational Approach for the Prediction of HIV Resistance Based on Amino Acid and Nucleotide Descriptors. *Molecules.* 2018;23(11):2751. doi 10.3390/molecules23112751
5. Paremskaia A.I. et al. Web Service for HIV Drug Resistance Prediction Based on Analysis of Amino Acid Substitutions in Main Drug Targets. *Viruses.* 2023;15(11):2245. doi 10.3390/v15112245
6. Tarasova O. et al. Identification of Molecular Mechanisms Involved in Viral Infection Progression Based on Text Mining: Case Study for HIV Infection. *Int J Mol Sci.* 2023;24(2):1465. doi 10.3390/ijms24021465

4 Симпозиум «Эволюционная, популяционная и медицинская геномика/генетика человека: компьютерные и экспериментальные подходы»

Symposium “Evolutionary, population and medical genomics/genetics of human: computational and experimental approaches”

doi 10.18699/bgrs2024-4.1-01

Генетическая структура ненцев по 21 аутосомному STR-маркеру

Валихова Л.В.*, Вагайцева К.В., Колесников Н.А., Харьков В.Н., Бочарова А.В., Степанов В.А.

Научно-исследовательский институт медицинской генетики, Томский НИМЦ РАН, Томск, Россия
** larisa.valikhova@medgenetics.ru*

Ключевые слова: ненцы; популяционная генетика; аутосомные STR-маркеры

Мотивация и цель: Ненцы – самодийский народ, населяющий территорию евразийского побережья Северного Ледовитого океана от Кольского полуострова до Таймыра. Современные ненцы делятся на две группы: лесных и тундровых [1]. Две группы ненцев значительно различаются друг от друга по языку (наличием двух диалектов), понимание между ними затруднено, а также имеются различия по антропологии [2]. Численность лесных ненцев в несколько раз меньше, чем тундровых. Внутри субэтноса тундровых ненцев выделяются роды, которые, в свою очередь, объединяются во фратрии, одна из которых, Харючи, включает роды самодийского происхождения, а фратрия Вануйто – роды аборигенного генетического субстрата, а также энецкого и хантыйского происхождения [3].
При изучении генофонда тундровых ненцев с помощью YSTR-маркеров было выяснено, что Y-хромосомный генофонд ненцев структурирован по родовому принципу. Выявлены значительные различия между двумя ненецкими фратриями [4].
В данной работе представлен результат анализа структуры генофонда ненцев по 21 аутосомному STR-маркеру, что дополнит имеющиеся данные об их структуре на основе других видов маркерных систем.
Методы и алгоритмы: Материалом исследования послужили выборки мужчин лесных ненцев, тундровых ненцев двух фратрий и разных родов. Всего было использовано 368 образцов, из них к лесным ненцам принадлежат 47, к фратирии Харючи – 216, к фратрии Вануйто – 105 (всего тундровых ненцев – 321). Забор венозной крови у доноров проводили с соблюдением процедуры письменного информированного согласия на проведение исследования. На каждого донора составлялась анкета с его родословной, указанием этнической принадлежности и мест рождения предков. Генотипирование образцов по аутосомным STR-маркерам проводили методом мультиплексной полимеразной цепной реакции с последующим анализом продуктов амплификации на автоматическом генетическом анализаторе НАНОФОР-05 («Синтол», Россия). При анализе использовался 21 аутосомный маркер: D20S1082, D6S474, D14S1434, D4S2666, D1S1677, D11S4463, D9S1122, D2S1776, D17S974, D1S1627, D3S4529, D2S1360, D3S1744, D9S2157, D17S1301, D8S1132, Penta D, D21S2050, D10S2325, D7S1517, Penta E. Для генотипирования использовали реакционный микс и активатор производства компании «Гордиз» (Россия). Реакционная смесь включала: 2 мкл активатора, 2 мкл реакционного микса, 0.16 мкл смеси праймеров, 1 мкл ДНК, 4.84 мкл ddH_2O для доведения до объема 10 мкл.
Анализ молекулярной дисперсии (AMOVA) проводили с использованием программного средства Arlequin v. 3.5 [5].

Результаты: При сравнении выборок лесных и тундровых ненцев на межпопуляционные различия пришлось 0.74 %. При сравнении двух фратрий тундровых ненцев (Харючи и Вануйто) на долю межпопуляционных различий приходится лишь 0.42 %. Большее значение межпопуляционных различий между лесными и тундровыми ненцами может быть объяснено большей изоляцией между ними по сравнению с фратриями тундровых ненцев. Также проводилось сравнение родов внутри каждой фратрии, отдельной группой считался род, который включал от 10 человек, все остальные роды были объединены в категорию «другие». Во фратрии Харючи были выделены роды: Лапсуй (35), Ядне (27), Яптунай (21), Яр (42), Яндо (17) Худи (16), «другие» (58). Во фратрии Вануйто были выделены роды Вануйто (43), Салиндер (32) и «другие» (30). Сравнение родов во фратрии Харючи показало отрицательное значение в межродовой дифференциации (–0.89). Это может свидетельствовать о крайне близком генетическом сходстве индивидов внутри фратрии Харючи, что подтверждается и другим исследованием [4]. При сравнении родов фратрии Вануйто получено самая большое значение в межпопуляционной дифференциации – 3.52. Это может быть связано с этногенезом данной фратрии, включением хантыйских родов – Тибичи, Вэхо, Неркыхы, Няданги, Пурунгуй, Салиндер, энецких родов – Марьик, Аседа, Паровых, Тёр, Оковай, которые могут усиливать генетическую дифференциацию фратрии [3]. Данные о внутрипопуляционных различиях представлены в табл. 1.

Таблица 1. Сравнение разных популяционных выборок

Группа сравнения	Межпопуляционные различия	Внутрипопуляционные различия
Лесные и тундровые ненцы	0.74	99.26
Фратрии Харючи и Вануйто	0.42	99.58
Роды внутри фратрии Харючи	–0.89	100.21
Роды внутри фратрии Вануйто	3.52	96.48

Выводы: Представленные данные подтверждают процессы этногенеза ненцев, их взаимоотношения как между субэтносами, так и между фратриями, внутри отдельных фратрий.

Финансирование: Работа выполнена в рамках государственного задания Министерства науки и высшего образования № 122112800026-1.

Genetic structure of the Nenets based on 21 autosomal STR markers

Valikhova L.V.*, Vagaitseva K.V., Kolesnikov N.A., Kharkov V.N.,
Bocharova A.V., Stepanov V.A.

Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia
** larisa.valikhova@medgenetics.ru*

Key words: Nenets; population genetics; autosomal STR markers

Motivation and Aim: The Nenets are a Samoyed people inhabiting the territory of the Eurasian coast of the Arctic Ocean from the Kola Peninsula to Taimyr. Modern Nenets are divided into two groups: forest and tundra [1]. The two groups of Nenets differ significantly from each other in language (the presence of two dialects), understanding between them is difficult, and there are also differences in anthropology [2]. The number of forest Nenets is several times smaller than that of the tundra Nenets. Within the subethnos of the Tundra Nenets, clans are distinguished, which in turn are united in phratries, one of which "Kharyuchi" – includes clans of Samoyed origin and the phratry "Vanuito" – includes clans of aboriginal genetic substrate, as well as Enets and Khanty origin [3].

When studying the gene pool of the tundra Nenets using YSTR markers, it was found that the Y-chromosomal gene pool of the Nenets is structured according to the generic principle. Significant differences were identified between the two Nenets phratries [4].

This paper presents the result of an analysis of the structure of the Nenets gene pool using 21 autosomal STR markers, which will complement the existing data on their structure based on other types of marker systems.

Methods and Algorithms: The research material consisted of samples of forest Nenets men, tundra Nenets men of two phratries and different clans. A total of 368 samples were used, of which 47 belong to the forest Nenets, 216 to the Kharyuchi phraty, and 105 to the Vanuito phratry (in total, 321 tundra Nenets). Venous blood was collected from donors in compliance with the procedure for written informed consent for the study. For each donor, a questionnaire was compiled with his pedigree, indicating his ethnicity and place of birth of his ancestors. Genotyping of samples using autosomal STR markers was carried out using the multiplex polymerase chain reaction method, followed by analysis of amplification products on an automatic genetic analyzer NANOFOR-05 (Sinthol, Russia). The analysis used 21 autosomal markers: D20S1082, D6S474, D14S1434, D4S2666, D1S1677, D11S4463, D9S1122, D2S1776, D17S974, D1S1627, D3S4529, D2S1360, D3S1744, 57, D17S1301, D8S1132, Penta D, D21S2050, D10S2325, D7S1517, Penta E For genotyping, we used a reaction mix and an activator produced by Gordiz (Russia). The reaction mixture included: 2 µl. activator, 2 µl. reaction mix, 0.16 µl primer mixtures, 1 µl DNA, 4.84 µl ddH_2O to make up to 10 µl volume. Analysis of molecular variance (AMOVA) was performed using the Arlequin software tool v 3.5 [5].

Results: When comparing samples of forest and tundra Nenets, interpopulation differences accounted for 0.74 %. When comparing two phratries of the tundra Nenets (Kharyuchi and Vanuito), interpopulation differences account for only 0.42 %. The greater significance of interpopulation differences between the forest and tundra Nenets can be explained by the greater isolation between them than between the phratries of the tundra Nenets. A comparison of clans within each phratry was also carried out; a clan that included 10 or more people was considered a separate group; all other clans were combined into the "other" category. In the Haryuchi phratry, the following clans were identified: Lapsui (35), Yadne (27), Yaptunay (21), Yar (42), Yando (17), Khudi (16), "others" (58). In the Vanuito phratry, two clans were also distinguished: Vanuito (43), Salinder (32), "others" (30). A comparison of genera in the Haryuchi phratry showed a negative value in intergeneric differentiation (–0.89). This may indicate an extremely close genetic similarity of individuals within the Haryuchi phratry, which is confirmed by another study [4]. When comparing the clans of the Vanuito phratry, the highest value in interpopulation differentiation was obtained – 3.52. This may be due to the

ethnogenesis of this phratry, the inclusion of the Khanty clans – Tibichi, Vekho, Nerkyhy, Nyadangi, Purungui, Salinder, Entets clans – Maryik, Aseda, Parovykh, Ter, Okovai, which can enhance the genetic differentiation of this moiety [3]. Data on intrapopulation differences are presented in Table 1.

Table 1. Comparison of different population samples

Comparison group	Interpopulation differences	Intrapopulation differences
Forest and tundra Nenets	0.74	99.26
Phratries of Haryuchi and Vanuito	0.42	99.58
Clans within the Haryuchi phratry	–0.89	100.21
Clans within the Vanuito phratry	3.52	96.48

Conclusions: The data presented above confirm the processes of ethnogenesis of the Nenets, their relationships both between subethnic groups and between phratries, within individual phratries.

Funding: The work was carried out within the framework of the State assignment of the Ministry of Science and Higher Education No. 122112800026-1.

Сисок литературы/References

1. Vasiliev V.I., Tugolukov V.A. Ethnographic research in Taimyr. *Sov Ethnography*. 1960;5:128-141
2. Khomich L.V. Problems of ethnogenesis and ethnic history of the Nenets. L.: Nauka, 1976
3. Kvashin Yu.N. The main elements of the clan structure of the Gydan Nenets. *Vestnik Arheologii, Antropologii i Etnografii*. 2001;3:1-4
4. Kharkov V.N. et al. Reconstruction of the origin of the Gydan Nenets based on genetic analysis of their generic structure using a new set of YSTR markers. *Genetics*. 2021;57(12):1403-1414
5. Excoffier L., Lischer H.E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10(3):564-567. doi 10.1111/j.1755-0998.2010.02847.x

doi 10.18699/bgrs2024-4.1-02

Изучение древней ДНК населения Северного Кавказа бронзового и железного веков

Габидуллина Л.[1]*, Джаубермезов М.[1, 2], Екомасова Н.[1, 2], Чагаров О.[3, 4], Атабиев Б.[5], Кадиева А.[6], Кушнеревич А.[7], Тамбетс К.[7], Виллемс Р.[7], Хуснутдинова Э.[1, 2]

[1] *Уфимский университет науки и технологий, Уфа, Россия*

[2] *Институт биохимии и генетики УФИЦ РАН, Уфа, Россия*

[3] *НИИ и Музей антропологии МГУ, Москва, Россия*

[4] *Институт археологии РАН, Москва, Россия*

[5] *Институт археологии Кавказа, Нальчик, Россия*

[6] *Государственный исторический музей, Москва, Россия*

[7] *Институт геномики Тартуского университета, Тарту, Эстония*

* *liliya.gab@gmail.com*

Ключевые слова: древняя ДНК; палеогенетика; бронзовый век; железный век; Северный Кавказ; сарматы; кобанская культура

Мотивация и цель: Географической особенностью Северного Кавказа является его расположение на границе Европы и Передней Азии, что вместе с соседством с Причерноморьем и Поволжьем послужило формированию миграционных и торговых путей, связавших Ближний Восток и Европу в различные исторические периоды, что в свою очередь отразилось на этногенезе народов региона. Формирование генетической структуры народов Северного Кавказа в большей степени протекало до начала выделения этнических групп в период бронзового и железного веков. В бронзовом веке на территории Кавказа развивались несколько археологических культур: майкопская, кура-аракская, северокавказская и завершающая период кобанская, во времена которой население уже было знакомо с железом, что говорит о начале перехода от бронзового к железному веку [1, 2]. В период железного века происходили повсеместные миграции на территории степной и горно-долинной Евразии, что отразилось на генофондах популяций [3]. В связи со сложными процессами этногенеза популяций на территории Северного Кавказа имеется вопрос о предковых компонентах в их генофондах – преемственности от древних археологических культур региона. Для решения данной задачи проводится исследование древней ДНК культур бронзового и железного века с территории Северного Кавказа.

Методы и алгоритмы: Для исследования был отобран палеоматериал в виде зубов 28 представителей майкопской, кобанской, сарматской и гуннской археологических культур. Информация о принадлежности образцов к определенной культуре и о возрасте исследуемых образцов первоначально была предоставлена археологами. Для подтверждения принадлежности исследуемых образцов к изучаемому временному периоду было проведено датирование методом радиоуглеродного анализа, всего таким образом было исследовано 10 образцов. Древняя ДНК выделялась из зубов, с дальнейшим созданием ДНК библиотек для NGS секвенирования на платформе Illumina HiSeq. В результате биоинформатического анализа данных NGS секвенирования были вычислены

показатели содержания эндогенной ДНК (среднее 32.15 %), уровни контаминации (<3 % 23 образца, 4–17 % 5 образцов) и степени повреждений древней ДНК по C>T заменам (8–27 %). Определен генетический пол исследуемых образцов – 11 женщин и 17 мужчин, а также по результатам метода READ исключены родственники из дальнейшего анализа (1 образец). Были определены гаплогруппы митохондриальной ДНК и Y-хромосомы. Для анализа полногеномных данных проводились PCA и ADMIXTURE с использованием набора данных "1240K+HO" лаборатории Dr David Reich, при этом 5 образцов из исследуемой выборки были исключены из анализа из-за низкого количества определенных SNP (<10 000 SNP) [4].

Результаты: Гаплогруппы митохондриальной ДНК образцов исследуемых археологических культур бронзового и железного веков представлены в основном западноевразийскими линиями гаплогрупп H, I, J, T, U, X (82.1 %). Восточноевразийские линии представлены ветвями гаплогрупп A, C, M (17.9 %). Среди гаплогрупп Y-хромосомы образцов бронзового и железного века преобладают переднеазиатские линии гаплогрупп G2 и J Y-хромосомы, встречаются также единичные носители гаплогруппы R1b, ранее выявленной у представителей степных культур. Образцы всех исследуемых археологических культур бронзового и железного веков по данным PCA анализа кластеризуются в основном с современными популяциями Кавказа и Ближнего Востока. По результатам ADMIXTURE анализа кавказский компонент является основным у большинства исследуемых образцов. Данные, полученные методами PCA и ADMIXTURE, соответствуют определенным нами гаплогруппам однородительских маркеров.

Выводы: Полученные результаты демонстрируют высокие показатели ряда предковых компонентов в генофондах археологических культур Северного Кавказа, что отражает сложные пути этногенеза в регионе в период бронзового и железного веков.

Финансирование: Исследование выполнено в рамках государственного задания Министерства науки и высшего образования РФ (№ 075-03-2024-123/1).

Study of ancient DNA of Bronze and Iron Ages population of Northern Caucasus

Gabidullina L.[1]*, Dzhaubermezov M.[1, 2], Ekomasova N.[1, 2], Chagarov O.[3, 4], Atabiev B.[5], Kadieva A.[6], Kushniarevich A.[7], Tambets K.[7], Villems R.[7], Khusnutdinova E.[1, 2]

[1] *Ufa University of Science and Technology, Ufa, Russia*

[2] *Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre, RAS, Ufa, Russia*

[3] *Laboratory of Contextual Antropology at the Institute of Archaeology, RAS, Moscow, Russia*

[4] *Institute of Archeology, RAS, Moscow, Russia*

[5] *Ltd. Institute for Caucasus Archaeology, Nalchik, Russia*

[6] *State Historical Museum, Moscow, Russia*

[7] *Institute of Genomics, University of Tartu, Tartu, Estonia*

* *liliya.gab@gmail.com*

Key words: ancient DNA; paleogenetics; Bronze Age; Iron Age; North Caucasus; Sarmatians; Koban culture

Motivation and Aim: A geographic feature of the North Caucasus is its location on the border of Europe and Western Asia, which, together with its proximity to the Black Sea and Volga regions, served to form migration and trade routes that connected the Middle East and Europe in various historical periods, which in turn affected the ethnogenesis of the peoples of the region. The formation of the genetic structure of the North Caucasus peoples took place to a greater extent before the beginning of the separation of ethnic groups, during the Bronze and Iron Ages. In the Bronze Age, several archaeological cultures developed in the Caucasus: Maikop, Kura-Araks, North Caucasus and the Koban culture, which ended the period, during which the population was already familiar with iron, which indicates the beginning of the transition from the Bronze to the Iron Age [1, 2]. During the Iron Age, widespread migrations took place in the territory of steppe and mountain-valley Eurasia, which ultimately affected in the gene pools of populations [3]. In connection with the complex processes of ethnogenesis of populations in the North Caucasus, there is a question about the ancestral components in their gene pools – continuity from the ancient archaeological cultures of the region. To solve this problem, a study of ancient DNA of Bronze and Iron Age cultures from the territory of the North Caucasus is being carried out.
Methods and Algorithms: For the study, paleomaterial was selected in the form of teeth from 28 representatives of the Maikop, Koban, Sarmatian and Hunnic archaeological cultures. Information about the culture of the samples and the age of the samples was provided by archaeologists. To confirm that the samples belonged to the time period being studied, dating was carried out using radiocarbon analysis; a total of 10 samples were dated. Ancient DNA was isolated from teeth, with further creation of DNA libraries for NGS sequencing on the Illumina HiSeq platform. As a result of bioinformatics analysis of NGS sequencing data, indicators of endogenous DNA content (average 32.15 %), contamination levels (<3 % 23 samples, 4–17 % 5 samples) and the degree of damage to ancient DNA by C>T substitutions (8 %) were calculated – 27 %). The genetic sex of the studied samples was determined – 11 women and 17 men; and based on the results of the READ method relatives were excluded from further analysis (1 sample). Mitochondrial DNA and Y-chromosome haplogroups were determined. To analyze the genome-wide data, PCA and ADMIXTURE were performed using the "1240K+HO" dataset from Dr David Reich; 5 samples from the study sample were excluded from the analysis due to the low number of identified SNPs (<10,000 SNPs) [4].
Results: The mitochondrial DNA haplogroups of the studied archaeological cultures of the Bronze and Iron Ages samples are represented mainly by West Eurasian lines of haplogroups H, I, J, T, U, X (82.1 %). East Eurasian lineages are represented by branches of haplogroups A, C, M (17.9 %). Among the Y-chromosome haplogroups of the Bronze and Iron Age samples, the West Asian lines of the G2 and J Y-chromosome haplogroups predominate; there are also single carriers of the R1b haplogroup, previously identified among representatives of steppe cultures. According to PCA analysis, samples of all studied archaeological cultures of the Bronze and Iron Ages are clustered mainly with modern populations of the Caucasus and the Middle East. According to the results of ADMIXTURE analysis, the Caucasian component is the main one in the majority of the studied samples. The data obtained by the PCA and ADMIXTURE methods correspond to the haplogroups of uniparental markers we determined.

Conclusion: The results obtained demonstrate high levels of a number of ancestral components in the gene pools of archaeological cultures of the North Caucasus, which reflects the complex paths of ethnogenesis in the region during the Bronze and Iron Ages.
Funding: The study was carried out within the state assignment of the Ministry of Science and Higher Education of the Russian Federation (No. 075-03-2024-123/1).

Список литературы/References

1. Мартынов А.И. Археология. М.: Высшая школа, 2005
   [Martynov A.I. Archeology. Moscow: Vysshaya Shkola, 2005 (in Russian)]
2. Археология. Эпоха бронзы Кавказа и Средней Азии. Ранняя и средняя бронза Кавказа. Москва: Наука, 1994
   [Archeology. Bronze Age of the Caucasus and Central Asia. Early and Middle Bronze Ages of the Caucasus. Moscow: Nauka, 1994 (in Russian)]
3. Степи европейской части СССР в скифо-сарматское время. М.: Наука, 1989
   [Steppes of the European part of the USSR in Scythian-Sarmatian times. Moscow: Nauka, 1989 (in Russian)]
4. Mallick S., Micco A., Mah M. et al. The allen ancient dna resource (AADR) a curated compendium of ancient human genomes. *Sci Data.* 2024;11:182. doi 10.1038/s41597-024-03031-7

doi 10.18699/bgrs2024-4.1-03

Изучение древней ДНК населения Северного Кавказа в 4000 до н.э. – 1800 н.э.

Джаубермезов М.[1, 2]*, Габидуллина Л.[1], Екомасова Н.[1, 2], Чагаров О.[3], Атабиев Б.[4], Коробов Д.[5], Науменко В.[6], Крутоголовенко К.[7], Кадиева А.[8], Прокофьев Р.[9], Вольная Г.[10], Кушнеревич А.[11], Тамбетс К.[11], Виллемс Р.[11], Хуснутдинова Э.[1, 2]

[1] *Уфимский университет науки и технологий, Уфа, Россия*

[2] *Институт биохимии и генетики УФИЦ РАН, Уфа, Россия*

[3] *НИИ и Музей антропологии МГУ, Москва, Россия*

[4] *Институт археологии Кавказа, Нальчик, Россия*

[5] *Институт археологии РАН, Москва, Россия*

[6] *Крымский федеральный университет им. В.И. Вернадского, Симферополь, Россия*

[7] *Всероссийское общество охраны памятников истории и культуры, Краснодар, Россия*

[8] *Государственный исторический музей, Москва, Россия*

[9] *ООО «Ростовская археологическая экспедиция», Ростов-на-Дону, Россия*

[10] *Институт истории и археологии, Владикавказ, Россия*

[11] *Институт геномики Тартуского университета, Тарту, Эстония*

* *murat-kbr@mail.ru*

Ключевые слова: древняя ДНК; мтДНК; популяции Северного Кавказа; аланы; кобанская культура

Мотивация и цель: Кавказ – преимущественно горный регион на границе Восточной Европы и Ближнего Востока. Растянувшись на более чем 1000 км между Черным и Каспийским морями, он тем не менее всегда оставался полупроницаемым для движущихся потоков древних культур [1, 2]. Располагаясь на стыке Европы и Азии, территория Кавказа на протяжении тысячелетий была связующим звеном между культурами Месопотамии, Ирана, Малой Азии, Средиземноморья и Восточной Европы [3], что приводило к взаимному влиянию как на культурный облик, так и на генофонд населения [4]. Учитывая особую ценность изучения археогеномики древнего населения данного региона, следует отметить, что до сих пор не освещены длительные исторические периоды истории Кавказа. В особенности это касается эпохи позднего бронзового века (конец II – начало I тыс. до н.э.) и отрезка от раннего Средневековья до периода Нового времени (V в. н.э. – начало XX в. н.э.).

Методы и алгоритмы: ДНК была извлечена из археологических материалов (зубы, височная кость, слуховые косточки) 100 человек с последующим проведением полногеномного секвенирования с низкой глубиной охвата и анализом их в контексте древних и современных генетических вариаций как изучаемого, так и близлежащих регионов. Покрытие геномов составило 0.0001–0.300X (медианное покрытие 0.0911X, содержание эндогенной ДНК 30.16 %).

Результаты: Радиоуглеродный анализ проведен для 31 из 100 исследованных образцов. За исключением единичных образцов, радиоуглеродное датирование практически соответствовало археологическому. Палитра гаплогрупп мтДНК как населения поздней бронзы, так и Средневековья, состоящая из 15 различных гаплогрупп (A, C, D, F, H, HV, I, J, K, M, N, T, U, W, X), схожа с разнообразием гаплогрупп мтДНК современных автохтонных популяций Кавказа. Разнообразие

гаплогрупп Y-хромосомы алан, состоящих из 6 различных гаплогрупп (E, G, I, J, Q и R), схоже с палитрой современных автохтонных популяций Северного Кавказа только на поздней стадии развития аланской культуры. Вызывает интерес, что гаплогруппы Y-хромосомы средневековых балкарцев (у 10 из 11 индивидов определены линии гаплогруппы R) разительным образом отличаются от разнообразия гаплогрупп, выявленных у средневековых осетин (гаплогруппа G определена у 6 из 8 образцов), однако близки с таковым у степных культур, что может указывать на различные процессы этногенеза для соседних популяций.
PCA-анализ указывает на существенную генетическую однородность всего разнообразия представленных в работе культур, а также на их сходство с современными популяциями Кавказа. Лишь единичные индивиды выбиваются из «кавказского» кластера, демонстрируя примесь центральноазиатского, восточноазиатского или европейского компонентов.
Выводы: Полученные результаты подтверждают теорию кавказского происхождения кобанской культуры поздней бронзы. Предварительный анализ позволяет предположить существование тесных контактов сарматов и алан с населением Северного Кавказа.
Финансирование: Исследование выполнено в рамках государственного задания Министерства науки и высшего образования РФ (№ 075-03-2024-123/1).

Study of ancient DNA of the population of North Caucasus in 4000 BCE – 1800 CE

Dzhaubermezov M.[1, 2]*, Gabidullina L.[1], Ekomasova N.[1, 2], Chagarov O.[3], Atabiev B.[4], Korobov D.[5], Naumenko V.[6], Krutogolovenko K.[7], Kadieva A.[8], Prokofiev R.[9], Volnaya G.[10], Khusnutdinova E.[1, 2]

[1] *Ufa University of Science and Technology, Ufa, Russia*
[2] *Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre, RAS, Ufa, Russia*
[3] *Laboratory of Contextual Antropology at the Institute of Archaeology, RAS, Moscow, Russia*
[4] *Ltd. Institute for Caucasus Archaeology, Nalchik, Russia*
[5] *Institute of Archeology, RAS, Moscow, Russia*
[6] *Crimean Federal University, Simferopol*
[7] *All-Russian Society for the Preservation of Historical and Cultural Monuments, Krasnodar, Russia*
[8] *State Historical Museum, Moscow, Russia*
[9] *Rostov Archaeological Expedition LLC, Rostov-on-Don, Russia*
[10] *Institute of History and Archeology, Vladikavkaz, Russia*
[11] *Institute of Genomics, University of Tartu, Tartu, Estonia*
* *murat-kbr@mail.ru*

Key words: ancient DNA; mtDNA; North Caucasus populations; Alans; Koban culture

Motivation and Aim: The Caucasus is a predominantly mountainous region on the border of Eastern Europe and the Middle East. The Caucasus stretches for more than 1,000 km between the Black and Caspian seas, however, it has always remained semipermeable to the moving streams of ancient cultures [1, 2]. Located at the junction of Europe and Asia, the territory of the Caucasus for thousands of years was a connecting link between the

cultures of Mesopotamia, Iran, Asia Minor, the Mediterranean and Eastern Europe [3], whichled to mutual influence on both the cultural image and the gene pool of the population [4]. Considering the special value of studying the archaeogenomics of the ancient population of this region, it should be noted that long historical periods of the Caucasus history have not yet been covered. This is especially true of the late Bronze Age (late II – early I millennium BCE) and the period from the early Middle Ages to the modern period (5th century CE – early 20th century CE).

Methods and Algorithms: DNA was extracted from archaeological materials (teeth, temporal bone, auditory ossicles) of 100 individuals, followed by low-depth whole-genome sequencing and analyzed in the context of ancient and modern genetic variation in both the study area and nearby regions. Genome coverage was 0.0001–0.300X (median coverage 0.0911X and endogenous DNA content 30.16 %).

Results: Radiocarbon dating was carried out for 31 of the 100 samples examined. With the exception of isolated samples, radiocarbon dating was almost identical to archaeological dating. The palette of mtDNA haplogroups of both Late Bronze and Middle Ages populations, consisting of 15 different haplogroups (A, C, D, F, H, HV, I, J, K, M, N, T, U, W, X), is similar with the diversity of mtDNA haplogroups of modern Caucasus autochthonous populations. The diversity of Y-chromosome haplogroups of the Alans, consisting of 6 different haplogroups (E, G, I, J, Q and R), became similar to the palette of modern autochthonous populations of the North Caucasus only at the late development stage of the Alan culture. It is of interest that the Y chromosome haplogroups of the medieval Balkars (in 10 out of 11 individuals, lines of haplogroup R were identified) are strikingly different from the diversity of Y chromosome haplogroups identified in medieval Ossetians (haplogroup G was identified in 6 out of 8 samples), however, they are still close to that of steppe cultures, that may indicate different processes of ethnogenesis for neighboring populations. PCA analysis indicates significant genetic homogeneity of the entire diversity of cultures presented in the work, as well as their similarity with modern Caucasus populations. Only a few individuals stand out from the “Caucasian” cluster, demonstrating an admixture of Central Asian, East Asian or European components.

Conclusion: The results obtained confirm the theory of the Caucasian origin of the Koban culture of the Late Bronze Age. Preliminary analysis suggests the existence of close contacts between the Sarmatians and Alans with the population of the North Caucasus.

Funding: The study is supported of the Ministry of Science and Higher Education of the Russian Federation (No. 075-03-2024-123/1).

Список литературы/References

1. Yunusbayev B., Metspalu M., Järve M. et al. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol Biol Evol*. 2012;29(1):359-365. doi 10.1093/molbev/msr221
2. Giemsch L. Hansen S. The Caucasus. Der Kaukasus Bridge between the urban centres in Mesopotamia and the Pontic steppes in the 4th and 3rd millennium BC Brücke zwischen den urbanen Zentren Mesopotamiens und der pontischen Steppe im 4. und 3. Jahrtausend v. Chr. Schriften des Archäologischen Museums, 2021
3. Археология. Эпоха бронзы Кавказа и Средней Азии. Ранняя и средняя бронза Кавказа. Москва: Наука, 1994.
   [Archeology. Bronze Age of the Caucasus and Central Asia. Early and Middle Bronze Ages of the Caucasus. Moscow: Nauka, 1994 (in Russian)]
4. Wang C.C., Reinhold S., Kalmykov A. et al. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat Commun*. 2019;10(1):590. doi 10.1038/s41467-018-08220-8

doi 10.18699/bgrs2024-4.1-05

Межпоколенная транформация популяционно-генетической структуры татар Сибири: изонимный подход

Имекина Д.О.[1]*, Ульянова М.В.[1], Тычинских З.А.[2], Лавряшина М.Б.[1]

[1] *Кемеровский государственный медицинский университет Минздрава России, Кемерово, Россия*

[2] *Тобольская комплексная научная станция УрО РАН, Тобольск, Россия*

* *dolinina_1993@mail.ru*

Ключевые слова: татары Сибири; генофонд; фамильный фонд; популяция

Мотивация и цель: Использование фамилий в качестве квазигенетического маркера хорошо зарекомендовало себя в исследованиях генетико-демографических характеристик популяции. Изучение фонда фамилий позволяет проанализировать популяционно-генетические показатели различных территориальных групп и их изменчивость на больших выборках, а также оценить межпоколенную трансформацию генетической структуры отдельных субэтнических групп.

Татары Сибири – собирательное название этнотерриториальных локальных групп татар, проживающих на территории Тюменской, Омской, Томской и Новосибирской областей. Широкий ареал расселения, высокая численность и генетическая гетерогенность локальных групп определяют необходимость комплексного исследования особенностей популяционно-генетической структуры их генофонда.

Целью настоящего исследования стало изучение и сопоставление фамильного состава в трех поколениях локальных групп татар Сибири для оценки трансформации популяционной структуры и контактов различных групп друг с другом.

Методы и алгоритмы: Источником информации для формирования баз данных о фондах фамилий татар в местах расселения этнографических групп послужили книги похозяйственного учета сельского населения. Обследованием охвачены территории девяти сельских поселений, входящих в ареал искеро-тобольских, иштякско-токузских, саргатско-утузских и татар-бухарцев. Фамилии собирались тотально за три временных отрезка 1950-е, 1980-е и 2010-е гг., что соответствует трем поколениям. Суммарный объем обследованных групп татар составил 18 630 человек с фондом фамилий более 1150 вариантов. На основе частот фамилий проанализированы генетические дистанции между поколениями внутри отдельных этнографических групп для оценки процессов трансформации фонда фамилий за три поколения. Проведен анализ генетических расстояний между исследованными этнографическими группами с целью выявления контактов – обмена фамилиями через возможные брачные миграции, а также динамики этого процесса. Генетические расстояния рассчитывали методом Нея.

Результаты: Исследование показало, что фамильный состав обследованных групп обладает своеобразием, особенно в части наиболее распространенных фамильных вариантов. Общие фамилии, встречающиеся во всех рассмотренных

группах татар, фиксируются редко и варьируют в исследованном временном интервале в диапазоне от 2.2 до 2.6 % в общей структуре фамилий. Анализ генетических расстояний между поколениями изученных этнографических групп татар Западной Сибири показал, что за три поколения (1952–2019 гг.) наибольшая трансформация фонда фамилий произошла у татар саргатско-утузской и иштякско-токузской групп – между поколениями 1950-х и 2010-х гг. Наибольшее постоянство фамильного состава отмечено у искеро-тобольских татар.
Выводы: Результаты осуществленного исследования свидетельствуют об эффективности фамилий татар Сибири как инструмента исследования структуры их популяций не только для фиксации трансформаций, но и для понимания их причин. В целом выявлена специфика демографического развития локальных групп татар Сибири, которая хорошо согласуется с географией расположения их этнических ареалов в пределах контактных зон и подтверждает перспективность применения фамилий для решения задач популяционной генетики.

Intergenerational transformation of the population-genetic structure of the tatars of Siberia: an isonymic approach

Imekina D.O.[1]*, Ulyanova M.V.[1], Tychinskikh Z.A.[2], Lavryashina M.B.[1]
[1] *Kemerovo State Medical University of the Ministry of Health of Russia, Kemerovo, Russia*
[2] *Tobolsk Complex Scientific Station, UB RAS, Tobolsk, Russia*
* *dolinina_1993@mail.ru*

Key words: Tatars of Siberia; gene pool; family fund; population

Motivation and Aim: The use of surnames as a quasi-genetic marker has proven itself well in studies of the genetic and demographic characteristics of the population. The study of the fund of surnames makes it possible to analyze the population-genetic indicators of various territorial groups and their variability in large samples, as well as to assess the intergenerational transformation of the genetic structure of individual subethnic groups.
Tatars of Siberia is the collective name of ethnoterritorial local groups of Tatars living in the Tyumen, Omsk, Tomsk and Novosibirsk regions. The wide area of settlement, the high number and genetic heterogeneity of local groups determine the need for a comprehensive study of the features of the population-genetic structure of their gene pool.
The purpose of this study was to study and compare the family composition in three generations of local groups of Siberian Tatars in order to assess the transformation of the population structure and contacts of various groups with each other.
Methods and Algorithms: The source of information for the formation of databases on the funds of Tatar surnames in the places of settlement of ethnographic groups were the books of household accounting of the rural population. The survey covered the territories of nine rural settlements included in the area of Isker-Tobolsk, Ishtyak-Tokuz, Sargat-Utuz and Tatar-Bukharians. Surnames were collected totally over three time periods of the 1950s, 1980s and 2010s, which corresponds to three generations. The total volume of the surveyed groups of Tatars amounted to 18,630 people with a fund of surnames of more than 1,150 variants. Based on the frequencies of surnames, the genetic distances between generations within individual ethnographic groups are analyzed to assess the

processes of transformation of the family name fund over three generations. The analysis of genetic distances between the studied ethnographic groups was also carried out in order to identify contacts – the exchange of surnames through possible marital migrations, as well as the dynamics of this process. Genetic distances were calculated using the Ney method.
Results: The study showed that the family composition of the surveyed groups has a peculiarity, especially in terms of the most common family variants. Common surnames found in all the studied groups of Tatars are rarely recorded and vary in the studied time interval in the range from 2.2 to 2.6 % in the general structure of surnames. An analysis of the genetic distances between generations of the studied ethnographic groups of Tatars of Western Siberia showed that over three generations (1952–2019), the greatest transformation of the family name fund occurred among the Tatars of the Sargat-Utuz and Ishtyak-Tokuz groups – between the generations of the 1950s and 2010s. The greatest constancy of the family composition was noted among the Isker-Tobolsk Tatars.
Conclusion: The results of the conducted research indicate the effectiveness of the surnames of the Tatars of Siberia as a tool for studying the structure of their populations not only for fixing transformations, but also for understanding their causes. In general, the study revealed the specifics of the demographic development of local groups of Siberian Tatars, which is in good agreement with the geography of their ethnic areas within contact zones and confirms the prospects of using surnames to solve problems of population genetics.

doi 10.18699/bgrs2024-4.1-06

Диахронный палеогенетический анализ средневекового населения юга Западной Сибири

Пилипенко А.[1]*, Трапезов Р.[1], Томилин М.[1], Черданцев С.[1], Пилипенко И.[1], Поздняков Д.[2], Молодин В.[2], Нестерова М.[2], Журавлев А.[1], Рыкун М.[3]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Институт археологии и этнографии СО РАН, Новосибирск, Россия*

[3] *Томский государственный университет, Томск, Россия*

* *alexpil@bionet.nsc.ru*

Ключевые слова: Западная Сибирь; средневековье; генетическая история населения; палеогенетический подход; диахронный анализ; митохондриальная ДНК; Y-хромосома

Мотивация и цель: В южных и центральных районах Западной Сибири на протяжении нескольких последних тысячелетий протекали этногенетические и культурогенетические процессы, конечным результатом которых стало формирование разнообразных групп коренного населения региона, включая представителей современных тюркоязычных, угорских и самодийских народов. Наиболее перспективным направлением исследования популяционно-генетических аспектов этих процессов является проведение диахронного исследования разновременных древних популяций региона методами палеогенетики. К настоящему времени накоплен значительный объем данных о генетической структуре популяций региона эпохи неолита, бронзового и раннего железного века. Однако средневековые популяции Западной Сибири остаются практически полностью неисследованными методами палеогенетики, хотя их анализ имеет важнейшее значение для реконструкции завершающих этапов формирования генетической структуры коренного населения региона. Целью нашей работы является реконструкция популяционно-генетических аспектов этногенетических процессов в Западной Сибири с помощью палеогенетического исследования диахронной модели, включающей репрезентативные серии образцов от всех основных групп средневекового населения южных районов Западной Сибири (а также выборки из популяций нового времени) с фокусом на анализ разнообразия митохондриальной ДНК и Y-хромосомы.

Методы и алгоритмы: В рамках данного исследования сформирована диахронная выборка общей численностью более 1000 палеоантропологических образцов (индивидов). В состав выборки вошли материалы из более 100 археологических памятников (могильников) с территории Барабинской лесостепи, Верхнего и Среднего Приобья, которые хронологически охватывают последние полторы тысячи лет, включая периоды раннего, развитого и позднего средневековья, а также новое время. Для этой выборки были получены образцы ДНК, оценена ее сохранность и проведен анализ структуры вариантов митохондриальной ДНК (для всех образцов с удовлетворительной сохранностью ДНК) и Y-хромосомы (для индивидов мужского пола с высокой сохранностью ДНК в останках). Разновременные популяции сравнивали между собой, а также с предшествующими популяциями региона и современными группами для выявления динамики генетического состава населения во времени.

Палеогенетические результаты сопоставляли с данными археологии и физической антропологии для получения комплексной реконструкции этногенетических процессов, протекавших на территории Западной Сибири на протяжении последних 1500 лет.

Результаты: Получены данные об особенностях генетического состава (на уровне разнообразия вариантов митохондриальной ДНК и Y-хромосомы) популяций, населявших Барабинскую лесостепь, Верхнее и Среднее Приобье на протяжении раннего, развитого и позднего средневековья, а также в новое время. Диахронный анализ позволил определить компоненты генофонда, которые маркируют генетическую преемственность разновременных средневековых популяций между собой и с предшествующими группами населения региона, а также выявить генетические компоненты, которые маркируют изменения генетического состава населения в результате миграционных процессов и этнокультурного взаимодействия. Предварительно идентифицированы компоненты, маркирующие ранние этапы тюркизации раннесредневекового населения лесостепной зоны Западной Сибири. Определена специфика преимущественно тюркоязычных, угорских и самодийских (по данным археологии) популяций, а также групп смешанного происхождения. Определены направления и географическая локализация зон основных межпопуляционных контактов в различные периоды средневековья. Сформирован обширный банк данных, включающий подробные археологические, антропологические и палеогенетические характеристики материалов, включенных в состав исследуемой диахронной модели.

Выводы: На протяжении различных этапов средневековья на территории южных районов Западной Сибири происходит интенсивное взаимодействие культурно- и генетически контрастных групп населения. Одним из основных направлений генетических потоков в регионе в этом период было взаимодействие по направлению север–юг (и в обратном направлении). Наиболее интенсивно это взаимодействие наблюдается на территории Верхнего Приобья, которое выполняет роль «артерии» для миграционных потоков в регионе. Именно здесь происходит наибольшее перемешивание генетических компонентов различного происхождения. Наше исследование завершает процесс формирования обширной диахронной палеогенетической модели, которая охватывает все основные популяции южных районов Западной Сибири за последние ~ 10 тысяч лет.

Финансирование: Исследование выполнено при финансовой поддержке Российского научного фонда, проект № 23-18-00834.

Diachronic paleogenetic analysis of medieval populations from the South of Western Siberia

Pilipenko A.[1]*, Trapezov R.[1], Tomilin M.[1], Cherdantsev S.[1], Pilipenko I.[1], Pozdnyakov D.[2], Molodin V.[2], Nesterova M.[2], Zhuravlev A.[1], Rykun M.[3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Institute of Archaeology and Ethnography, SB RAS, Novosibirsk, Russia*

[3] *Tomsk State University, Tomsk, Russia*

* *alexpil@bionet.nsc.ru*

Key words: Western Siberia; the Middle Ages; genetic history of the population; paleogenetic approach; diachronic analysis; mitochondrial DNA; Y-chromosome

Motivation and Aim: As a result of ethnogenetic and culturogenetic processes. which took place in the southern and central regions of Western Siberia during the past few millennia, various groups of the regional indigenous population have formed, including representatives of modern Turkic-speaking, Ugric and Samoyed peoples. The most promising direction of studying the population-genetic aspects of these processes is to conduct a diachronic study of regional ancient populations by paleogenetic methods. A significant amount of data on the genetic structure of the Neolithic, Bronze and Early Iron Age populations from Western Siberia was accumulated. However, the medieval populations of Western Siberia remain almost completely unstudied by paleogenetic methods, although their analysis is of crucial importance for reconstructing the final stages of the genetic structure formation for indigenous populations of the region. The aim of our work is to reconstruct the population-genetic aspects of ethnogenetic processes in Western Siberia using a paleogenetic study of a diachronic model that includes representative series of samples from all major groups of the medieval West Siberian populations with a focus on analyzing the variability of mitochondrial DNA and Y-chromosome.

Methods and Algorithms: Within the framework of this study, a diachronic sample of more than 1,000 paleoanthropological samples (individuals) was formed. The sample includes materials from more than 100 archaeological sites (burial grounds) from the territory of the Baraba forest-steppe, Upper and Middle Ob region, which chronologically cover the last fifteen hundred years, including Middle Ages, as well as modern times. For these samples we obtained a total DNA extracts, analyzed the DNA preservation level and obtained data on the structure of mitochondrial DNA lineages (for all samples with sufficient DNA preservation) and the Y-chromosome (for male individuals with high DNA preservation in the remains). The populations of different medieval periods were compared with each other, as well as with the previous populations of the region and modern groups to identify the chronological dynamics of the genetic composition. The paleogenetic results were compared with data from archaeology and physical anthropology to obtain a comprehensive reconstruction of the ethnogenetic processes that took place in Western Siberia over the past 1,500 years.

Results: Data on the features of the genetic composition (at the level of diversity of mitochondrial DNA and Y-chromosome variants) of populations from the Baraba forest-steppe, Upper and Middle Ob region dated by the different parts of the Middle Ages, as well as by modern times, have been obtained. Diachronic analysis made it possible to identify the components of the gene pool that mark the genetic succession of different medieval populations, as well as with previous regional groups, as well as to identify genetic components that mark changes in the genetic composition of the population as a result of migration processes and ethnocultural interaction. The components marking the early stages of Turkification of the early medieval population of the West Siberian forest-steppe zone have been preliminarily identified. The specifics of predominantly Turkic-speaking, Ugric and Samoyed (according to archaeological data) populations, as well as groups of mixed origin, were determined. The directions and geographical localization of main zones of interpopulation contacts in different Middle Ages periods were determined. An extensive data bank has been formed, including detailed archaeological,

anthropological and paleogenetic characteristics of the materials included in the diachronic model under study.

Conclusion: During different stages of the Middle Ages, intensive interaction of culturally and genetically contrasting population groups took place in the southern regions of Western Siberia. One of the main directions of genetic flows in the region in this period was the interaction in the north-south direction (and in the opposite direction). This interaction is most intensively observed in the territory of the Upper Ob region, which acts as an "artery" for migration flows in the region. The most intensive interaction of genetic components of various origins takes place in this territory. Our study completes the process of forming an representative diachronic paleogenetic model that covers all the main populations of the southern regions of Western Siberia over the past ~ 10 thousand years.

Funding: The study was supported by the Russian Science Foundation, project No. 23-18-00834.

doi 10.18699/bgrs2024-4.1-07

Изучение ассоциаций признаков агрессивного поведения с рядом генетических локусов на выборке из представителей современной армянской популяции

Прошаков П.[1]*, Ревякина Л.[1], Лазебный О.[1], Бутовская М.[2]

[1] *Институт биологии развития им. Н.К. Кольцова РАН, Москва, Россия*

[2] *Институт этнологии и антропологии им. Н.Н. Миклухо-Маклая РАН, Москва, Россия*

* *ub3dco@gmail.com*

Ключевые слова: гены-кандидаты; агрессивное поведение; моноэтничная выборка

Мотивация и цель: Агрессивное поведение как количественный признак имеет очень сложную структуру, и контроль со стороны генетических детерминант происходит на фоне определенной этнической и культурной среды [1]. Для более объективного генетического анализа агрессивного поведения необходимо учитывать изменчивость, связанную с этническим происхождением. Ассоциации между полиморфизмом генов и поведенческими признаками оценивали с использованием генов, связанных с функционированием нервной системы, таких как дофаминовая и серотониновая системы, а также генов, связанных с половыми гормонами. Например, такие гены, как *OXTR*, *HTR2A* и *AR*, известны важной ролью в регуляции агрессивного поведения [2–6]. Связь SNP-локуса rs53576 гена *OXTR* с агрессией может быть объяснена влиянием на скорость трансляции белка [7], тогда как SNP-локус rs6311 гена *HTR2A* может влиять на экспрессию гена [8]. Полиморфный микросателлитный локус *AR*$(CAG)_n$ в гене *AR* также имеет значение, влияя на конформацию и активность белка [9]. Окситоцин, важный гормон, регулирует психику и эмоции у млекопитающих [10]. Исследования показывают его связь с агрессией [6], но данные о генотипе *GG* и эмпатии противоречивы [11]. Локус rs6311 рассматривается как маркер агрессивного поведения [12]. Меньшее количество CAG-повторов в микросателлитном локусе *AR* ассоциируется с агрессивными чертами [13].

Большинство исследований проводилось на традиционных этнических группах, что ограничивает понимание влияния генетических факторов на поведение в современной европейской культуре. Наше исследование направлено на подтверждение влияния трех перечисленных локусов на агрессивное поведение на примере моноэтничной выборки армянских студентов.

Методы и алгоритмы: Исследование проводилось на выборке студентов Ереванского государственного университета в период с 2014 по 2016 гг. Генотипирование проводилось на 231 человеке для локусов *OXTR* rs53576 и *HTR2A* rs6311, а для локуса *AR*$(CAG)_n$ – только на 229 индивидуумах. Оценка агрессивного поведения производилась с помощью опросников Басса–Перри и по реактивной-проактивной агрессии [14, 15]. Генотипирование проводилось методами ПЦР-ПДРФ для SNP-локусов и фрагментным анализом для локуса *AR*$(CAG)_n$. Статистическая обработка данных выполнена с использованием программ GenAlEx v 6.51 [16], Statistica v.10 (StatSoft, Inc., США), Structure v.2.3.4 [17], APSampler v. 3.61 [18]. Различия считались значимыми при $p < 0.05$, с учетом

поправки Бенжамини–Хохберга на множественные сравнения [19]. При анализе влияния локуса *AR* на признаки агрессии набор аллелей был предварительно ранжирован на «короткие» и «длинные», характеризующиеся 12–21 повтором CAG и 22–29 повторами соответственно.

Результаты: В локусе *OXTR* rs53576 преобладает аллель *G* (0.72) с генотипом *GG* у 52 % образцов. В *HTR2A* rs6311 аллели *A* и *G* представлены примерно равно (0.48 и 0.52 соответственно), с генотипом *AG* у 46 % образцов. Распределение генотипов соответствует равновесию, подтвержденному тестом χ^2. Выявлено 17 аллелей локуса *AR*$(CAG)_n$, преобладают аллели с 20, 21 и 23 повторами (суммарная частота 0.49). Кластерный анализ (Structure) показал однородность выборки – наличие одного кластера.

Были рассчитаны средние, стандартные отклонения, максимальные и минимальные значения признаков агрессии, оцененных опросниками Басса–Перри и реактивной-проактивной агрессии. Распределения шкал агрессии в основном не отличаются от нормального, за исключением физической агрессии у женщин и вербальной и проактивной агрессии в объединенной выборке. Тест Стьюдента показал значимые различия между мужчинами и женщинами по нескольким шкалам агрессии: физическая агрессия (p = 0.000000028), враждебность (p = 0.0327), проактивная (p = 0.000000028) и реактивная агрессия (p = 0.0151), которые выше у мужчин. Значения альфа Кронбаха для всех шкал составляют 0.7892, для опросника Басса–Перри – 0.7164, для реактивной-проактивной агрессии – 0.7329.

Проведен ассоциативный анализ как по общим данным, так и отдельно по мужчинам и женщинам. Четырехфакторный дисперсионный анализ подтвердил влияние пола на уровень физической, проактивной и реактивной агрессии, а также показал обратную зависимость враждебности между мужчинами и женщинами.

Трехфакторный дисперсионный анализ выявил влияние полиморфизма локуса *AR*$(CAG)_n$ на физическую агрессию у мужчин и локуса *OXTR* rs53576 на вербальную агрессию у женщин. Также обнаружено влияние взаимодействия локусов *OXTR* rs53576 и *AR*$(CAG)_n$ на различные шкалы агрессии у женщин и ассоциация взаимодействия *OXTR* rs53576 × *HTR2A* rs6311 с реактивной агрессией у женщин.

Анализ в программе APSampler выявил сочетания генотипов, характерные как для более, так и для менее агрессивной части выборки, как для всех, так и для отдельных групп мужчин и женщин. Для более агрессивных представителей: *OXTR* rs53576 – *AG* + *HTR2A* rs6311 – *A*. Для менее агрессивных представителей: *OXTR* rs53576 – *GG* + *AR*$(CAG)_n$ – 24.

Выводы: Исследовано влияние генетических факторов на агрессивное поведение у мужчин и женщин, подтверждены гендерные различия [20], а также влияние локусов *OXTR* rs53576 и *AR*$(CAG)_n$ на гнев и агрессию. Дальнейшие исследования необходимы для проверки воспроизводимости результатов и уточнения механизмов влияния [21].

У мужчин обнаружен эффект генотипа локуса *AR*$(CAG)_n$ на физическую агрессию, согласующийся с [13]. Однако авторы [22] подвергли этот эффект сомнению. Важность этнического фактора подтверждается [5]. У женщин выявлен независимый эффект локуса *OXTR* rs53576 на вербальную агрессию, противоречащий [11, 23], но согласующийся с нашими данными [5]. Эмпатия и ответ на стресс связаны с этим эффектом [10, 11, 24].

Различия в результатах могут быть обусловлены половой спецификой экспрессии генов у мужчин и женщин, что затрудняет обнаружение отдельных эффектов локусов *OXTR* rs53576 и *AR*$(CAG)_n$ на агрессию. Взаимодействие этих локусов подтверждено как у мужчин на вербальную агрессию, так и у женщин на вербальную, проактивную и реактивную агрессию. Механизм взаимодействия остается неясным из-за различий в физиологических системах, к которым относятся данные локусы.

У женщин наблюдается влияние взаимодействия локусов *OXTR* rs53576 и *HTR2A* rs6311 на реактивную агрессию. Гомозиготные носители локуса *HTR2A* rs6311 с аллелем *A* локуса *OXTR* rs53576 проявляют пониженный уровень агрессии, в то время как гетерозиготные носители генотипа *AG* имеют максимальный уровень. Сочетание генотипа *GG* локуса *OXTR* rs53576 с любым генотипом локуса *HTR2A* rs6311 характеризуется средним уровнем агрессии.

Наблюдаются сложные взаимосвязи между аллелями локуса *HTR2A* rs6311 и между локусами генов *HTR2A* и *OXTR*. Генотип *GG* локуса *HTR2A* rs6311 ассоциируется с повышенным уровнем суицидального поведения и агрессии [12]. Ранее был показан эффект взаимодействия этнической принадлежности и генотипа локуса *HTR2A* rs6311 на агрессию.

На втором этапе, с применением непараметрического ассоциативного анализа, мы выявили комбинации генотипов для двух частей выборки: *OXTR* rs53576 – *GG* + *AR*$(CAG)_n$ – 24 для «менее агрессивных» и *OXTR* rs53576 – *AG* + *HTR2A* rs6311 – *A* для «более агрессивных», что соответствует результатам дисперсионного анализа. Мы подтверждаем влияние локуса *AR*$(CAG)_n$ на уровень физической агрессии у мужчин и локуса *OXTR* rs53576 на уровень вербальной агрессии у женщин в нашей выборке. Все обнаруженные эффекты, особенно взаимодействия локусов, требуют дальнейшей проверки и верификации на выборках из различных культурно-этнических групп, возможно, с привлечением более крупных выборок и применением более мощных методов анализа, таких как GWAS (полногеномный поиск ассоциаций).

Финансирование: Исследование поддержано в рамках раздела государственного задания ИБР РАН 2024 года № 0088-2024-0011.

The study of associations between traits of aggressive behavior and a set of genetic loci in a sample of representatives from the contemporary Armenian population

Proshakov P.[1]*, Revyakina L.[1], Lazebny O.[1], Butovskaya M.[2]

[1] *Koltzov Institute of Developmental Biology, RAS, Moscow, Russia*

[2] *Miklukho-Maklai Institute of Ethnology and Anthropology, RAS, Moscow, Russia*

* *ub3dco@gmail.com*

Key words: candidate genes; aggressive behavior; monoethnic sample

Motivation and Aim: Aggressive behavior, as a quantitative trait, has a very complex structure, and genetic control occurs against the background of specific ethnic and cultural environments [1]. For a more objective genetic analysis of aggressive behavior, it is necessary to consider variability associated with ethnic origin. Associations between

gene polymorphisms and behavioral traits were assessed using genes related to the functioning of the nervous system, such as the dopamine and serotonin systems, as well as genes associated with sex hormones. Genes such as *OXTR*, *HTR2A*, and *AR* are known to play an important role in regulating aggressive behavior [2–6]. The association of the SNP locus rs53576 in the *OXTR* gene with aggression can be explained by its influence on the rate of protein translation [7], while the SNP locus rs6311 in the *HTR2A* may affect gene expression [8]. The polymorphic microsatellite locus *AR*$(CAG)_n$ in the *AR* gene also plays a significant role by influencing protein conformation and activity [9].

Oxytocin, an important hormone, regulates the psyche and emotions in mammals [10]. Research indicates its association with aggression [6], but data regarding the *GG* genotype and empathy are contradictory [11]. The rs6311 locus is considered as a marker for aggressive behavior [12]. A lower number of CAG repeats in the *AR* microsatellite locus is associated with aggressive traits [13].

Most studies have been conducted on traditional ethnic groups, limiting the understanding of the influence of genetic factors on behavior in modern European culture. Our study aims to confirm the influence of the studied loci on aggressive behavior using a mono-ethnic sample of Armenian students.

Methods and Algorithms: The study was conducted on a sample of students from Yerevan State University from 2014 to 2016. Genotyping was performed on 231 individuals for the *OXTR* rs53576 and *HTR2A* rs6311 loci, and only on 229 individuals for the *AR*$(CAG)_n$ locus. Assessment of aggressive behavior was carried out using the Buss-Perry questionnaires and measures of reactive-proactive aggression [14, 15]. Genotyping was conducted using PCR-RFLP methods for SNP loci and fragment analysis for the *AR*$(CAG)_n$ locus. Statistical analysis was performed using GenAlEx v 6.51 [16], Statistica v.10 (StatSoft, Inc., USA), Structure v.2.3.4 [17], and APSampler v. 3.61 [18]. Differences were considered significant at $p < 0.05$, with adjustments made for multiple comparisons using the Benjamini–Hochberg correction [19]. When analyzing the influence of the *AR* locus on aggression traits, the allele set was pre-ranked into 'short' and 'long' categories, characterized by 12–21 and 22–29 CAG repeats, respectively.

Results: In the *OXTR* rs53576 locus, the *G* allele predominates (0.72), with the *GG* genotype present in 52 % of the samples. In *HTR2A* rs6311, alleles *A* and *G* are represented approximately equally (0.48 and 0.52, respectively), with the *AG* genotype in 46 % of the samples. The distribution of genotypes conforms to equilibrium, confirmed by the χ^2 test. Seventeen alleles of the *AR*$(CAG)_n$ locus were identified, with alleles containing 20, 21, and 23 repeats predominating (combined frequency of 0.49). Cluster analysis (Structure) demonstrated sample homogeneity, indicating the presence of a single cluster.

Means, standard deviations, maximum, and minimum values of aggression traits assessed by the Buss–Perry questionnaire and reactive-proactive aggression were calculated. The distributions of aggression scales mostly did not deviate from normal, except for physical aggression in women and verbal and proactive aggression in the combined sample. The Student's t-test revealed significant differences between men and women across several aggression scales: physical aggression ($p = 0.000000028$), hostility ($p = 0.0327$), proactive ($p = 0.000000028$), and reactive aggression ($p = 0.0151$), all of which were higher in men. The Cronbach's alpha values for all scales are 0.7892, for the Buss–Perry questionnaire – 0.7164, and for reactive-proactive aggression – 0.7329.

Associative analysis was conducted using separate aggression scales from both questionnaires. An associative analysis was performed using both overall data and separately for men and women. A four-way analysis of variance confirmed the influence of gender on the level of physical, proactive, and reactive aggression, and also showed a reverse relationship of hostility between men and women.

Three-way analysis of variance revealed the influence of polymorphism of the *AR*(CAG)$_n$ locus on physical aggression in men and the *OXTR* rs53576 locus on verbal aggression in women. Additionally, the interaction between the *OXTR* rs53576 and *AR*(CAG)$_n$ loci was found to influence various aggression scales in women, along with an association of the interaction *OXTR* rs53576 × *HTR2A* rs6311 with reactive aggression in women.

The analysis in the APSampler software revealed combinations of genotypes characteristic of both more and less aggressive segments of the sample, both overall and within individual groups of men and women. For more aggressive individuals: *OXTR* rs53576 – *AG* + *HTR2A* rs6311 – *A*. For less aggressive individuals: *OXTR* rs53576 – *GG* + *AR*(CAG)$_n$ – 24.

Conclusion: The influence of genetic factors on aggressive behavior in men and women has been investigated, confirming gender differences [20], as well as the influence of *OXTR* rs53576 and *AR*(CAG)$_n$ loci on anger and aggression. Further research is necessary to confirm the reproducibility of the results and to clarify the mechanisms of influence [21].

An effect of the *AR*(CAG)$_n$ locus on physical aggression in men has been found, consistent with [13]. However, [22] have cast doubt on this effect. The importance of ethnic factors is confirmed by [5]. An independent effect of the *OXTR* rs53576 locus on verbal aggression in women was identified, contradicting [11, 23] but consistent with our data [5]. Empathy and stress response are associated with this effect [10, 11, 24].

Differences in results may be due to gender-specific gene expression in men and women, complicating the detection of individual effects of *OXTR* rs53576 and *AR*(CAG)$_n$ loci on aggression. The interaction of these loci has been confirmed both in men for verbal aggression and in women for verbal, proactive, and reactive aggression. The mechanism of interaction remains unclear due to differences in physiological systems to which these loci belong.

In women, the interaction between *OXTR* rs53576 and *HTR2A* rs6311 loci has been observed to influence reactive aggression. Homozygous carriers of the *HTR2A* rs6311 locus with the *A* allele of the *OXTR* rs53576 locus exhibit lower levels of aggression, while heterozygous carriers of the *AG* genotype have the highest level. The combination of the *GG* genotype of the *OXTR* rs53576 locus with any genotype of the *HTR2A* rs6311 locus is characterized by an intermediate level of aggression.

Complex relationships are observed between alleles of the *HTR2A* rs6311 locus and between loci of the *HTR2A* and *OXTR* genes. The *GG* genotype of the *HTR2A* rs6311 locus is associated with increased levels of suicidal behavior and aggression [12]. Previously, an interaction effect of ethnic origin and genotype of the *HTR2A* rs6311 locus on aggression has been demonstrated.

In the second stage, using non-parametric associative analysis, we identified genotype combinations for two segments of the sample: *OXTR* rs53576 – *GG* + *AR*(CAG)$_n$ – 24 for “less aggressive” and *OXTR* rs53576 – *AG* + *HTR2A* rs6311 – *A* for “more aggressive”, consistent with the results of the analysis of variance.

We confirm the influence of the *AR*$(CAG)_n$ locus on the level of physical aggression in men and the *OXTR* rs53576 locus on the level of verbal aggression in women in our sample. All detected effects, especially the interactions of loci, require further verification and validation in samples from different cultural-ethnic groups, possibly involving larger samples and the use of more powerful analysis methods, such as GWAS (genome-wide association study).

Funding: The study is supported by Government program of basic research in Koltzov Institute of Developmental Biology of the Russian Academy of Sciences in 2024 No. 0088-2024-0011.

Список литературы/References

1. Алфимова М.В., Трубников В.И. Психогенетика агрессивности. *Вопросы психологии*. 2000;(6):112-121
   [Alfimova M.V., Trubnikov V.I. Psychogenetics of Aggressiveness. *Voprosy Psikhologii = Issues of Psychology*. 2000;(6):112-121 (in Russian)]
2. Давыдова Ю.Д., Литвинов С.С., Еникеева Р.Ф., Малых С.Б., Хуснутдинова Э.К. Современные представления о генетике агрессивного поведения. *Вавиловский журнал генетики и селекции*. 2018;22(6):716-725
   [Davydova J.D., Litvinov S.S., Enikeeva R.F., Malykh S.B., Khusnutdinova E.K. Recent advances in genetics of aggressive behavior. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2018;22(6):716-725 (in Russian)]
3. Ковш Е.М., Воробьева Е.В., Ермаков П.Н. Обзор современных исследований психогенетических факторов агрессивного поведения. *Российский психологический журнал*. 2014;11(4):91-103
   [Kovsh E.M., Vorobyeva E.V., Ermakov P.N. Review of current research on psychogenetic factors of aggressive behavior. *Rossiyskiy psikhologicheskiy zhurnal = Russian Psychological Journal*. 2014;11(4):91-103 (in Russian)]
4. Васильев В.А. Молекулярная психогенетика: исследования девиантного агрессивного поведения человека. *Генетика*. 2011;47(9):1157-1168
   [Vasilyev V.A. Molecular psychogenetics of deviant aggressive behavior in humans. *Russ J Genet*. 2011;47(9):1157-1168 (in Russian)]
5. Butovskaya M.L., Butovskaya P.R., Vasilyev V.A. et al. Serotonergic gene polymorphisms (5-HTTLPR, 5HTR1A, 5HTR2A), and population differences in aggression: traditional (Hadza and Datoga) and industrial (Russians) populations compared. *J Physiol Anthropol*. 2018;37(1)10. doi 10.1186/s40101-018-0171-0
6. Butovskaya M., Burkova V., Vasilyev V. et al. Fertility and infant survival in men and women from rural regions of Northern Tanzania: gene candidates and sex-specific genetic associations. *J Anthropol Sci*. 2020;98. doi 10.4436/JASS.98018
7. Hartl D.L., Moriyama E.N., Sawyer S.A. Selection intensity for codon bias. *Genetics*. 1994;138(1):227-234. doi 10.1093/genetics/138.1.227
8. Falkenberg V.R., Gurbaxani B.M., Unger E.R., Rajeevan M.S. Functional genomics of serotonin receptor 2A (HTR2A): interaction of polymorphism, methylation, expression and disease association. *Neuromol Med*. 2011;13:66-76. doi 10.1007/s12017-010-8138-2
9. Gelmann E.P. Molecular biology of the androgen receptor. *J Clin Oncol*. 2002;20(13):3001-3015. doi 10.1200/JCO.2002.10.018
10. Carter C.S., Grippo A.J., Pournajafi-Nazarloo H., Ruscio M.G., Porges S.W. Oxytocin, vasopressin and sociality. *Prog Brain Res*. 2008;170:331-336. doi 10.1016/S0079-6123(08)00427-5
11. Rodrigues S.M., Saslow L.R., Garcia N., John O.P., Keltner D. Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. *Proc Natl Acad Sci USA*. 2009;106(50):21437-21441. doi 10.1073/pnas.0909579106
12. Giegling I., Hartmann A.M., Möller H.-J., Rujescu D. Anger- and aggression-related traits are associated with polymorphisms in the 5-HT-2A gene. *J Affective Disord*. 2006;96(1-2):75-81. doi 10.1016/j.jad.2006.05.016
13. Butovskaya M.L., Vasilyev V.A., Lazebny O.E. et al. Aggression and polymorphisms in AR, DAT1, DRD2, and COMT genes in Datoga pastoralists of Tanzania. *Sci Rep*. 2013;3:3148. doi 10.1038/srep03148

14. Buss A.H., Perry M. The aggression questionnaire. *J Personal Social Psychol.* 1992;63(3):452-459. doi 10.1037/0022-3514.63.3.452
15. Raine A., Dodge K., Loeber R. et al. The reactive-proactive aggression questionnaire: differential correlates of reactive and proactive aggression in adolescent boys. *Aggress Behav*. 2006;32(2):159-171. doi 10.1002/ab.20115
16. Peakall R., Smouse P.E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*. 2012;28(19):2537-2539. doi 10.1093/bioinformatics/bts460
17. Pritchard J.K., Stephens M., Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945-959. doi 10.1093/genetics/155.2.945
18. Favorov A.V., Andreewski T.V., Sudomoina M.A., Favorova O.O., Parmigiani G., Ochs M.F. A Markov chain Monte Carlo technique for identification of combinations of allelic variants underlying complex diseases in humans. *Genetics*. 2005;171(4):2113-2121. doi 10.1534/genetics.105.048090
19. Benjamini Y., Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Soc Ser B (Methodol)*. 1995;57(1):289-300
20. Ильин Е.П. Психология агрессивного поведения. СПб.: Питер, 2014
[Ilyin E.P. Psikhologiya aggressivnogo povedeniya = The Psychology of Aggressive Behavior. SPb.: Piter, 2014 (in Russian)]
21. Cupaioli F.A., Zucca F.A., Caporale C., Lesch K.P., Passamonti L., Zecca L. The neurobiology of human aggressive behavior: Neuroimaging, genetic, and neurochemical aspects. *Prog Neuropsychopharmacol Biol Psychiatry*. 2021;106:110059. doi 10.1016/j.pnpbp.2020.110059
22. Valenzuela N.T., Ruiz-Pérez I., Rodríguez-Sickert C., Polo P., Muñoz-Reyes J.A., Yeste-Lizán A., Pita M. The relationship between androgen receptor gene polymorphism, aggression and social status in young men and women. *Behav Sci (Basel).* 2022;12(2):42. doi 10.3390/bs12020042
23. Parris M.S., Grunebaum M.F., Galfalvy H.C. et al. Attempted suicide and oxytocin-related gene polymorphisms. *J Affect Disord*. 2018;238:62-68. doi 10.1016/j.jad.2018.05.022
24. Басова А.Г. Теоретические аспекты взаимосвязи эмпатии и агрессии. *Молодой ученый*. 2012;8(43):256-258
[Basova A. G. Theoretical aspects of the relationship between empathy and aggression. *Molodoy Ucheny = Young Scientist*. 2012;8(43):256-258 (in Russian)]

doi 10.18699/bgrs2024-4.1-08

Генотипирование аллелей *ε2, ε3, ε4* гена *APOE* в выборках коренных популяций Восточной Сибири

Тийс Р.[1]*, Табиханова Л.[1], Личман Д.[1], Воронина Е.[2], Осипова Л.[1], Филипенко М.[2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

** kruosana@mail.ru*

Ключевые слова: аполипопротеин-Е; липиды; ген *APOE*; *rs7412 C>T* и *rs429358 T>C*; *ε2, ε3, ε4*; болезнь Альцгеймера; сердечно-сосудистые и онкологические заболевания; старение; коренные народы Сибири

Мотивация и цель: Аполипопротеин-Е участвует в метаболизме липидов, витамина D, в процессах старения, а также во многих других важных биологических механизмах в организме человека [1]. Полиморфизм в гене *APOE* имеет огромное влияние на риск развития и тяжесть течения многих серьезных заболеваний, включая болезнь Альцгеймера, гиперпротеинемию III типа, ишемическую болезнь сердца, атеросклероз, онкологические и многие другие заболевания [2–5]. Поэтому важность исследования этого полиморфизма в человеческих популяциях не вызывает сомнений. Однако для некоторых коренных популяций такие данные скудны или вовсе отсутствуют. В связи с этим целью нашего исследования явилось изучение распределения частот полиморфных вариантов гена *APOE rs7412C>T* (*Arg158Cys*) и *rs429358T>C* (*Cys112Arg*), приводящих к образованию трех аллелей, *ε2, ε3, ε4*, которые, в свою очередь, дают шесть *APOE*-генотипов: *ε2/ε2*, *ε3/ε3*, *ε4/ε4*, *ε2/ε3*, *ε2/ε4*, *ε3/ε4*, в популяциях коренных народов, проживающих на территориях Восточной Сибири (якутов Нюрбинского (N = 102) и Усть-Алданского (N = 99) улусов, восточных (N = 132) и западных (N = 280) бурятов и русских (N = 121)), а также проведение межпопуляционных сравнений между изученными нами и некоторыми мировыми популяциями.

Методы и алгоритмы: Для выделения геномной ДНК из лейкоцитарной фракции применялся стандартный метод фенол-хлороформной экстракции. Генотипирование однонуклеотидных полиморфизмов *rs7412* и *rs429358* проводилось методом Real-time ПЦР с использованием конкурирующих TaqMan-зондов, комплементарных полиморфным участкам ДНК. Структуру праймеров и зондов подбирали по последовательностям, доступным в базе данных NCBI (http://www.ncbi.nlm.nih.gov/), с использованием программ UGENE (version 1.14, http://ugene.unipro.ru/) и OligoAnalyzer (version 1.0.3, https:// eu.idtdna.com/pages/tools/oligoanalyzer). Популяционные частоты полиморфных вариантов вычисляли на основе наблюдаемых частот генотипов. Соответствие эмпирически наблюдаемого распределения частот генотипов теоретически ожидаемому распределению, равновесному по закону Харди–Вайнберга, проверяли с использованием критерия Пирсона ($\chi 2$; при p (H–W) > 0.05 равновесие выполняется, где p (H–W) – значение вероятности отклонения от равновесного распределения Харди–Вайнберга). Оценку достоверности различий в частотах

полиморфных вариантов между исследованными выборками проводили по критерию $\chi 2$ с применением поправки Йейтса на непрерывность; при $p < 0.025$ (с поправкой на множественность сравнения: 0.025 = 0.05/2) результаты считали статистически значимыми.

Результаты: Полученные нами частоты *APOE*-генотипов и *ε2*, *ε3*, *ε4* аллелей в изученных выборках коренных народов приведены в табл. 1 и 2 соответственно.

Таблица 1. Распределение частот *APOE*-генотипов в исследованных популяционных выборках

Генотипы	Популяции/Частоты генотипов, %				
	Якуты, Нюрб. улус	Якуты, Усть-Алд. улус	Восточные буряты	Западные буряты	Русские
ε2/ε2	0	1 (1)	0	0	0.8 (1)
ε2/ε3	16.7 (17)*	13.1 (13)	7.6 (10)	3.2 (9)	11.6 (14)
ε2/ε4	0.9 (1)	4 (4)	1.5 (2)	1.4 (4)	0.8 (1)
ε3/ε3	65.7 (67)	63.6 (63)	59.1 (78)	61.4 (172)	62 (75)
ε3/ε4	16.7 (17)	16.2 (16)	25 (33)	31.1 (87)	23.1 (28)
ε4/ε4	0	2 (2)	6.8 (9)	2.9 (8)	1.7 (2)
Всего чел.	102	99	132	280	121

* Количество человек в выборке с соответствующим генотипом.

Таблица 2. Частоты аллелей *ε2*, *ε3*, *ε4* гена *APOE* в популяциях коренных народов Восточной Сибири

Популяция	Частота *ε2*, %	Частота *ε3*, %	Частота *ε4*, %
Якуты, Нюрбинский улус (N=102)	8.8	82.4	8.8
Якуты, Усть-Алданский улус (N=99)	9.6	78.3	12.1
Буряты восточные (N=132)	4.5	75.4	20.1
Буряты западные (N=280)	2.3	78.6	19.1
Русские (N=121)	7.0	79.4	13.6

Наше исследование показало, что среди пяти изученных популяций значения частот полиморфных вариантов гена *APOE* заметно варьируют для *ε2* и *ε4*. При этом по частотам *ε2* заметно различаются две выборки якутов, имеющих наибольшие значения среди остальных популяций, что достоверно отличает якутов и западных бурят. Частоты варианта *ε4* имеют наивысшие значения среди восточных и западных бурят, что значимо отличает эти выборки от якутов. По частотам варианта *ε3* наблюдается наименьший разброс между популяциями. Отдельно стоит сказать о русских Восточной Сибири. Полученные частоты всех трех изученных полиморфных варианта гена *APOE* в этой выборке схожи и не имеют достоверных отличий от таковых для других европеоидов. Среди остальных изученных выборок коренных народов от русских достоверно отличаются западные буряты в частотах варианта *ε2* и западные и восточные буряты в частотах *ε4*. По *ε3* значимых различий между русскими и коренными народами не наблюдается.

Выводы: Если применить понятие «нормального» и «дисфункционального» полиморфных вариантов *APOE* гена, то совместная распространенность дисфункциональных *ε4* и *ε3*, или, условно говоря, генетический груз, в процентном содержании довольно высок во всех пяти изученных нами популяциях. В связи с этим можно предполагать, что в этих популяциях вероятен риск развития тех заболеваний, с которым ассоциирован полиморфизм в гене

APOE, особенно это может касаться сердечно-сосудистых заболеваний (атеросклероза, ожирения, ИБС) в случае, если коренные народы «отходят» от своего привычного уклада жизни, от привычного способа питания в пользу употребления пищи, богатой быстрыми углеводами и/или насыщенными жирами. Если брать во внимание адаптивное значение *ε4*, в том плане, что его носители лучше усваивают витамин D из пищи [6], то можно предположить, что высокие значения частот этого варианта среди западных и восточных бурят, живущих в Восточной Сибири, по сравнению с другими коренными народами нашего исследования, также имеют адаптивное значение для данных популяций, проживающих в условиях резко континентального климата. В то же время вариант *ε4* связан с более высоким уровнем холестерина в крови у его носителей, а в условиях Восточной Сибири, когда требуются высокие энергозатраты для лиц, ведущих преимущественно сельский образ жизни, наличие такого варианта может быть адаптивным. Однако, чтобы подтвердить все наши предположения о влиянии полиморфизма гена *APOE* на изученные нами популяции, необходимо увеличить количество выборок коренных жителей, живущих в других климатических условиях, а также провести рандомизированное исследование среди групп больных и здоровых лиц. Тем не менее полученные на данный момент результаты важны и могут применяться в интересах персонализированной медицины, исследования истории расселения коренных популяций по территории Сибири, процессов их адаптации к суровым условиям среды.

Финансирование: Исследование выполнено в рамках государственного задания ИЦиГ СО РАН (№ FWNR-2022-0021).

Genotyping of the *ε2*, *ε3*, *ε4* alleles of the *APOE* gene in samples of indigenous populations of Eastern Siberia

Tiis R.[1]*, Tabikhanova L.[1], Lichman D.[1], Voronina E.[2], Osipova L.[1], Filipenko M.[2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

* *kruosana@mail.ru*

Key words: apolipoprotein-E; lipids; *APOE* gene; *rs7412 C > T* and *rs429358T > C*; *ε2*, *ε3*, *ε4*; Alzheimer's disease; cardiovascular and oncological diseases; aging; indigenous peoples of Siberia

Motivation and Aim: Apolipoprotein-E is involved in lipid metabolism, vitamin D metabolism, aging, and many other important biological mechanisms in the human body [1]. Polymorphism in the *APOE* gene has a huge impact on the risk of developing and severity of the course of many serious diseases, including Alzheimer's disease, type III hyperproteinemia, coronary heart disease, atherosclerosis, cancer and many other diseases [2–5]. Therefore, the importance of investigating this polymorphism in human populations is unquestionable. However, these data are scarce or absent for some indigenous populations. Therefore, the aim of our study was to investigate the frequency distribution of polymorphic variants of the *APOE* gene *rs7412C>T* (*Arg158Cys*) and *rs429358T>C* (*Cys112Arg*) resulting in three alleles, *ε2*, *ε3*, *ε4*, which in turn yield six *APOE*-genotypes: *ε2/ε2*, *ε3/ε3*, *ε4/ε4*, *ε2/ε3*, *ε2/ε4*, *ε3/ε4*, in populations of indigenous peoples living in the territories of Eastern Siberia (Yakuts of Nyurbinsky (N=102) and Ust-Aldansky (N=99) Uluses, Eastern (N=132) and Western (N=280) Buryats and

Russians (N = 121)), as well as interpopulation comparisons between the populations we studied and some world populations.

Results: The frequencies of *APOE*-genotypes and *ε2*, *ε3*, *ε4* alleles in the studied indigenous samples are shown in Tables 1 and 2, respectively.

Table 1. Frequency distribution of *APOE* genotypes in the studied population samples

Genotypes	Populations/Genotype frequencies, %				
	Yakuts, Nyurba	Yakuts, Ust-Aldan	Eastern Buryats	Western Buryats	Russians
ε2/ε2	0	1 (1)	0	0	0.8 (1)
ε2/ε3	16.7 (17)*	13.1 (13)	7.6 (10)	3.2 (9)	11.6 (14)
ε2/ε4	0.9 (1)	4 (4)	1.5 (2)	1.4 (4)	0.8 (1)
ε3/ε3	65.7 (67)	63.6 (63)	59.1 (78)	61.4 (172)	62 (75)
ε3/ε4	16.7 (17)	16.2 (16)	25 (33)	31.1 (87)	23.1 (28)
ε4/ε4	0	2 (2)	6.8 (9)	2.9 (8)	1.7 (2)
Total persons	102	99	132	280	121

* Number of people in the sample with the corresponding genotype.

Table 2. Allele frequencies of *ε2*, *ε3*, *ε4* alleles of the *APOE* gene in the populations of indigenous peoples of Eastern Siberia

Population	Frequency of *ε2*, %	Frequency of *ε3*, %	Frequency of *ε4*, %
Yakuts, Nyurbinsky ulus(N=102)	8.8	82.4	8.8
Yakuts, Ust-Aldansky ulus(N=99)	9.6	78.3	12.1
Eastern Buryats (N=132)	4.5	75.4	20.1
Western Buryats (N=280)	2.3	78.6	19.1
Russians (N=121)	7.0	79.4	13.6

Our study showed that among the 5 populations studied, the frequencies of polymorphic variants of the *APOE* gene vary markedly for *ε2* and *ε4*. The frequencies of *ε2* differ markedly between the two samples of Yakuts, which have the highest values among the other populations, which reliably distinguishes Yakuts and Western Buryats. The frequencies of the *ε4* variant have the highest values among Eastern and Western Buryats, which significantly distinguishes these samples from Yakuts. The frequencies of the *ε3* variant show the smallest variation between the populations. It is worth mentioning the Russians of Eastern Siberia separately. The obtained frequencies of all three studied polymorphic variants of the *APOE* gene in this sample are similar and have no significant differences from those of other Caucasians. Among the other studied samples of indigenous peoples, Western Buryats reliably differ from Russians in the frequencies of the variant *ε2* and Western and Eastern Buryats in the frequencies of *ε4*; no significant differences between Russians and indigenous peoples are observed for *ε3*.

Conclusion: If we apply the concept of "normal" and "dysfunctional" polymorphic variants of *APOE* gene, then the joint prevalence of dysfunctional *ε4* and *ε3*, or conditionally speaking, genetic load, in percentage content is rather high in all 5 populations studied by us. In this regard, it can be assumed that in all these populations there is a probable risk of developing those diseases associated with the polymorphism in the *APOE* gene, especially cardiovascular diseases (atherosclerosis, obesity, CHD) if indigenous peoples "depart" from their usual way of life, from the usual way of eating in favor of eating food rich in fast carbohydrates and/or saturated fats. If we take into account the adaptive significance of *ε4*, in that its carriers better absorb vitamin D from food [6], we can assume that the high frequency values of this variant among Western and Eastern Buryats living in Eastern Siberia, compared to other indigenous peoples of

our study, also have an adaptive significance for these populations living in a sharply continental climate. At the same time, the ε4 variant is associated with higher blood cholesterol levels in its carriers, and in the conditions of Eastern Siberia, when high energy expenditures are required for individuals leading a predominantly rural lifestyle, the presence of such a variant may be adaptive. However, to confirm all our assumptions about the influence of the *APOE* gene polymorphism on the populations we studied, it is necessary to increase the number of samples of indigenous people living in other climatic conditions, as well as to conduct a randomized study among groups of sick and healthy individuals. Nevertheless, the results obtained so far are important and can be used in the interests of personalized medicine, the study of the history of settlement of indigenous populations in Siberia, and the processes of their adaptation to harsh environmental conditions.

Funding: the research was carried out within the framework of a state assignment of Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia (No. FWNR-2022-0021).

Список литературы/References

1. Mahley R.W., Rall S.C. Apolipoprotein E: far more than a lipid transport protein. *Annu Rev Genomics Hum Genet*. 2000;1:507-537
2. Mahley R.W., Weisgraber K.H., Huang Y. Apolipoprotein E4: a causative factor and therapeutic target in neuropathology, including Alzheimer's disease. *Proc Natl Acad Sci USA*. 2006;103:5644-5651
3. Григорьева И.Н., Нотова Т.Е. Полиморфизм гена аполипопротеина Е, желчнокаменная болезнь, сахарный диабет 2 типа и нарушения липидного обмена. *Атеросклероз*. 2023;19(1):47-56. doi 10.52727/2078-256X-2023-19-1-47-56
[Grigorieva I.N., Notova T.E. Polymorphism of the apolipoprotein E gene, gallstone disease, type 2 diabetes mellitus and lipid metabolism disorders. *Atherosclerosis*. 2023;19(1):47-56 (in Russian)]
4. Зуева И.Б., Улитина А.С., Гораб Д.Н., Москаленко М.В., Дубин М.В. Полиморфизм гена ApoE у пациентов с метаболическим синдромом и когнитивными расстройствами. *Артериальная гипертензия*. 2012;18(5):421-428
[Zueva I.B., Ulitina A.S., Gorab D.N., Moskalenko M.V., Dubin M.V. Polymorphism of the ApoE gene in patients with metabolic syndrome and cognitive disorders. *Arterial Hypertension*. 2012;18(5):421-428 (in Russian)]
5. Кох Н.В., Воронина Е.Н., Лифшиц Г.И. Полиморфизм гена *APOE*, как фактор риска дислипидемии, атеротромбоза и потенциальный фармакогенетический маркер терапии статинами у пациентов высокого риска сердечно-сосудистых осложнений. *Атеросклероз и дислипидемии*. 2016;3:107-115
[Kokh N.V., Voronina E.N., Lifshits G.I. *APOE* gene polymorphism is a risk factor for dyslipidemia and potential pharmacogenetic marker of lipid-lowering therapy. *Ateroskleroz i Dislipidemii*. 2016;3:107-115 (in Russian)]
6. Боровкова Н.П., Шереметьева В.А., Евсюков А.Н., Спицын В.А. Закономерности распределения аллелей аполипопротеина Е (*APOE*) среди мирового народонаселения. *Вестник Московского университета. Серия XXIII. Антропология*. 2010;2:21-35
[Borovkova N.P., Sheremetyeva V.A., Evsyukov A.N., Spitsyn V.A. Patterns of distribution of apolipoprotein E (*APOE*) alleles among the world population. *Moscow University Anthropology Bulletin*. 2010;2:21-35 (in Russian)]

doi 10.18699/bgrs2024-4.1-09

Генетическая структура сарматского населения Нижнего Поволжья и Южного Приуралья

Томилин М.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
tomilin@bionet.nsc.ru

Ключевые слова: сарматы; палеогенетика; древняя ДНК; митохондриальная ДНК; Y-хромосома; диахронный анализ

Мотивация и цель: Сарматы – конфедерация воинственных кочевых племен, которые с IV в. до н.э. по IV в. н.э. заселяли западную часть степного пояса Евразии от бассейна Дуная на западе до Уральских гор на востоке. Благодаря античным письменным источникам современные ученые имеют представление о ключевой роли сарматов в различных этнокультурных и политических событиях западной части Евразийских степей. Помимо античной литературы, важную роль сарматов подтверждают археологические исследования многочисленных памятников сарматской культуры. За долгую (более ста лет) историю исследования сарматского населения с точки зрения археологии и антропологии удалось выработать периодизацию их культуры, а также охарактеризовать сарматов с точки зрения физической антропологии. Однако, несмотря на высокую степень изученности, до сих пор остаются нерешенными вопросы их происхождения и связи с населением других регионов Евразии. Одним из наиболее эффективных подходов для объективной реконструкции истории сарматского населения является проведение палеогенетического исследования останков носителей различных этапов сарматской культуры. Получение данных о генетическом составе сарматских популяций, его формировании и динамике позволяет более детально реконструировать историю этих групп населения, игравших высокую роль в политических и этнокультурных процессах на западе степного пояса Евразии.

Цель работы – исследование генетической структуры савромато-сарматского населения Нижнего Поволжья и Приуралья по данным вариабельности митохондриальной ДНК и Y-хромосомы и реконструкция механизмов формирования генофонда сарматских популяций в контексте происходивших в этих регионах исторических и этнокультурных процессов.

Методы и алгоритмы: Экстракция древней ДНК из репрезентативных серий палеоантропологических образцов, анализ нуклеотидной последовательности гипервариабельного сегмента I и филогенетически значимых позиций в кодирующей части мтДНК, филогенетический и филогеографический анализ, определение половой принадлежности останков, анализ структуры Y-хромосомы с помощью анализа аллельных профилей STR-маркеров и статуса филогенетически значимых ОНП Y-хромосомы. Работы выполнены на базе специализированной инфраструктуры для палеогенетических исследований в межинститутской лаборатории молекулярной палеогенетики и палеогеномики ИЦиГ СО РАН. В процессе работы соблюдались все необходимые требования по предотвращению влияния контаминации и верификации полученных результатов.

Результаты:

- Получены серии образцов для предсавроматского и савроматского времени Нижнего Поволжья (N=16).
- Исследована репрезентативная выборка сарматских материалов Нижнего Поволжья, относящихся к разным этапам культуры (раннесарматскому (IV в. до н.э. – I в. до н.э.), среднесарматскому (I – первая половина II в. н.э.), позднесарматскому (вторая половина II–IV в. н.э.)) общей численностью более 100 образцов (индивидов).
- Получены результаты по разнообразию вариантов мтДНК раннесарматского населения Приуралья (N=15).
- Получены первые данные по разнообразию вариантов мтДНК разновременных групп населения Нижнего Поволжья эпохи бронзы (ямная, катакомбная, срубная культуры).
- Получены предварительные данные по мужскому генофонду (разнообразие вариантов Y-хромосомы) всех перечисленных групп древнего населения.

Выводы:

- Было установлено, что сарматы не демонстрируют преемственности относительно срубной культуры, обитавшей в Нижнем Поволжье в эпоху поздней бронзы, согласно данным по мтДНК и Y-хромосомы. Полученные данные свидетельствуют о высокой роли мигрантов в происхождении сарматского населения Нижнего Поволжья.
- Материалы из Приуралья демонстрируют признаки интенсивного миграционного притока в данный регион населения из более восточных районов степного пояса Евразии, включая южные районы Сибири.

Финансирование: Исследование поддержано за счет средств бюджетного проекта ИЦиГ СО РАН № FWNR-2022-0011.

Genetic structure of the Sarmatian people of the Lower Volga region and the Southern Ural region

Tomilin M.*

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

** tomilin@bionet.nsc.ru*

Key words: Sarmatians; paleogenetics; ancient DNA; mitochondrial DNA; Y-chromosome; diachronic analysis

Motivation and Aim: The Sarmatians were a confederation of warlike nomadic tribes that inhabited the western part of the Eurasian steppe from the Danube basin in the west to the Ural Mountains in the east since the 4th century BC to the 4th century AD. According to ancient written sources, modern scholars have an idea of the key role of the Sarmatians in various ethno-cultural and political events in the western part of the Eurasian steppes. In addition to ancient literature, the important role of the Sarmatians is also confirmed by archaeological studies of numerous monuments of the Sarmatian culture. During the long (more than a hundred years) history of the study of the Sarmatian population studied by archeology and anthropology, it was possible to develop a periodisation of their culture and to characterise the Sarmatians in terms of physical anthropology. However,

despite the high degree of study, the questions of their origin and connection with the population of other regions of Eurasia are still unresolved. One of the most effective approaches for the objective reconstruction of the history of the Sarmatian population is the palaeogenetic study of remains belonging to the different stages of the Sarmatian culture. Obtaining data about the genetic composition of Sarmatian populations, its formation and dynamics allows a more detailed reconstruction of the history of these populations, which played a high role in political and ethno-cultural processes in the western Eurasian steppe.

The aim of this work is to study the genetic structure of the Sauromatian-Sarmatian population of the Lower Volga region and the Ural region following the data of mitochondrial DNA and Y-chromosome variability and to reconstruct the mechanisms of gene pool formation of the Sarmatian populations in the context of historical and ethnocultural processes that took place in these regions.

Methods and Algorithms: Extraction of ancient DNA from representative series of paleoanthropological samples; nucleotide sequence analysis of the hypervariable segment I and phylogenetically significant positions in the mtDNA coding part, phylogenetic and phylogeographic analysis, sex determination of the remains, Y-chromosome structure analysis via analysis of STR marker allele profiles and status of phylogenetically significant Y-chromosome SNPs. The work was performed on the specialised infrastructure for palaeogenetic studies in the interinstitutional laboratory of molecular palaeogenetics and palaeogenomics of ICG SB RAS. In the process of work fulfilment, all necessary requirements to prevent the influence of contamination and verification of the obtained results were maintained.

Results:

- A series of specimens for the Pre-Sauromatian and Sauromatian periods of the Lower Volga region were obtained (N=16).
- A representative sample of Sarmatian materials from the Lower Volga region belonging to different stages of culture (Early Sarmatian (IV century BC – I century BC), Middle Sarmatian (I – first half of II century AD), Late Sarmatian (second half of II–IV century AD)) was studied, totalling more than 100 specimens (individuals).
- The results on the diversity of mtDNA variants of the Early Sarmatian population of the Ural region (N=15) were obtained.
- The first data on the mtDNA variant diversity of different groups in the Lower Volga region population of the Bronze Age (Yamnaya, Catacomb, and Shrubnaya cultures) were obtained.
- Preliminary data on the male gene pool (Y-chromosome variant diversity) of all mentioned ancient population groups were obtained.

Conclusion:

- It was found that the Sarmatians do not display continuity with respect to the Srubnaya culture that inhabited the Lower Volga region in the Late Bronze Age, according to mtDNA and Y-chromosome data. The obtained data indicate a high role of migrants in the origin of the Sarmatian population of the Lower Volga region.
- The Ural materials indicate evidence of intensive migration influx to this region from further eastern parts of the Eurasian steppe, including the southern parts of Siberia.

Funding: The study is supported from the budget project of ICG SB RAS No. FWNR-2022-0011.

doi 10.18699/bgrs2024-4.1-10

Филогения и филогеография Y-хромосомной гаплогруппы N1a1a1a1a-F1419

Харьков В.[1]*, Валихова Л.[1], Хитринская И.[1], Зарубин А.[1], Адамов Д.[2], Степанов В.[1]

[1] *Научно-исследовательский институт медицинской генетики, Томский НИМЦ РАН, Томск, Россия*

[2] *Северо-Восточный федеральный университет им. М.К. Аммосова, Якутск, Россия*

** vladimir-kharkov@medgenetics.ru*

Ключевые слова: Y-хромосома; гаплогруппа N1a1; филогения; филогеография

Мотивация и цель: Подробная детализация филогении гаплогрупп Y-хромосомы и оценки возраста различных линий за счет оценки результатов полногеномного секвенирования вносит существенный вклад в определение их территориального происхождения, историю демографических процессов и миграций. Обнаружение новых терминальных SNP для различных сублиний Y-хромосомы позволяет уточнить специфичность всех исследуемых мужских линий, детализировать их этнические различия и оценить время расхождения всех ветвей любой гаплогруппы. Целью работы явилась подробная филогения северо-евразийской гаплогруппы N1a1a1a1a-F1419 с помощью полногеномного секвенирования Y-хромосом из различных филогенетических ветвей основных линий и отдельных редких образцов со специфичными YSTR-гаплотипами. Для более детальной филогении этноспецифичных линий было увеличено количество образцов мужчин из новых популяционных выборок. Это позволяет максимально детально охарактеризовать структуру этой гаплогруппы на основании извлечения данных по новым YSNP из полных геномов, а также сопоставлении их с данными гаплотипов по YSTR-маркерам.

Методы и алгоритмы: Были извлечены специфичные для гаплогруппы N1a1a1a1a-F1419 SNP из файлов с полногеномным секвенированием методом NGS у 202 образцов ДНК из различных популяций Сибири, Восточной Европы, Дальнего Востока, Средней Азии и Северного Кавказа. Проведен расчет возраста отдельных ветвей по соответствующим выборкам образцов [1]. Рассчитан общий размер прочитанных нуклеотидных последовательностей образцов выборки внутри границ области combBED [2]. Отбор производных вариантов осуществлялся по координате референсной последовательности hg38, попадающей в пределы регионов combBED. Расчет погрешности оценки возраста основывается на предположении о пуассоновском характере процесса SNP мутаций [3]. Анализ STR-гаплотипов проводили с применением 45 YSTR (DYS19, 385a, 385b, 388, 389I, 389II, 390, 391, 392, 393, 426, 434, 435, 436, 437, 438, 439, 442, 444, 445, 448, 449, 456, 458, 460, 462, 481, 504, 505, 518, 525, 531, 533, 537, 552, 570, 576, 635, 643, YCAIIa, YCAIIb, GATA H4.1, Y-GATA-A10, GGAAT1B07). Полученные оценки возраста сравнивались с данными работы [4] и расчетами YFull.

Результаты: Детализированы филогения и территориальное распределение специфичных сублиний гаплогруппы N1a1a1a1a-F1419 Y-хромосомы за счет расширения спектра SNP. У многих образцов мужчин, относящихся к сибирским и европейским популяциям, обнаружено большое количество новых, не описанных ранее SNP. В общем массиве маркеров они составляют около 20 %, но для различных образцов из этноспецифичных ветвей их доля в терминальных SNP порой доходит до 100 %. Эти результаты позволяют более точно выявлять и дифференцировать ее варианты у мужчин различных родов и субэтнических групп. По результатам выявления новых YSNP и привязке образцов к детализированным сублиниям в различных гаплогруппах был проведен расчет возраста по стандартной формуле с учетом рассчитанного ранее темпа мутирования SNP, указанного в различных статьях и на сайте YFull. Проведено генотипирование широкого набора информативных для детализации этой гаплогруппы YSNP, извлеченных из просеквенированных образцов на имеющихся популяционных выборках. Почти все SNP полностью подтвердили свои аллели на уже просеквенированных образцах и на всех остальных образцах из этих гаплогрупп. Показано, что большая часть новых терминальных YSNP, в сочетании с YSTR-гаплотипами, обладает большой этнопопуляционной специфичностью. В большинстве случаев сравнительный анализ случайно выбранных образцов позволяет провести этнотерриториальную привязку образца с точностью до этноса, субэтноса или рода, а в отдельных случаях и до конкретного населенного пункта и носителей гаплотипов YSTR. Выполнено построение филогенетического древа этой гаплогруппы по 202 образцам и медианных сетей YSTR для ее различных линий от общего предка по этой гаплогруппе до отдельных образцов мужчин со специфичными терминальными YSNP. В качестве примера можно привести оценку возраста специфичной якутско-эвенкийской линии N-M1984, который составил 960 лет. Возраст более ранней линии N-M2016, к которой вместе якутами и эвенками принадлежат отдельные образцы узбеков и хакасов, определен в 2520 лет. Общий возраст гаплогруппы N1a1a по всем секвенированным образцам сибирских, среднеазиатских и европейских популяций составил 14 020 лет. Возраст специфичной линии для бурятов составил 980 лет, у сибирских татар – 1400 лет, у хакасов и шорцев – 740 лет. Практически для всех генотипированных образцов получена точная привязка к конкретным специфичным сублиниям, которые полностью соответствуют их этнической, субэтнической и родовой принадлежности. В качестве примера можно привести данные о различиях по Y-хромосомным гаплогрупам у тундровых и лесных ненцев. У лесных ненцев полностью доминирует гаплогруппа N1a1a1a1a2a1c1~ (Y13852), к которой принадлежат все мужчины рода Пяк. Возраст рода Пяк лесных ненцев составил 451 год, род тундровых ненцев Салиндер из фратрии Вануйто – 331 год, рода Тибичи – 216 лет, что хорошо согласуется с данными антропологов и этнографов. Предковая для коми, марийцев и удмуртов линия N-Z35179 имеет возраст 2800 лет, общая для марийцев и удмуртов европейская линия N-Z35198 – 1660 лет. Линия N-Y40117, к которой принадлежат три образца коми, имеет возраст 200 лет. Возраст линии N-Z35193 у трех удмуртов составляет 300 лет. В целом результаты расчетов возраста ветвей гаплогруппы N1a1a1a1a-F1419 согласуются с данными сравниваемых работ, но в некоторых случаях имеются различия. Часть наших образцов формирует отдельные сублинии, которые не были описаны ранее, и позволяет более подробно охарактеризовать филогению

этих линий. Была проведена картографическая привязка основных линий гаплогруппы N1a1a1a1a-F1419 по их частоте в исследованных популяциях.
Выводы: Филогения гаплогруппы N1a1a1a1a-F1419 подтверждает, что местом ее исходного распространения по территории Сибири и Европы была зона Южной Сибири или прилегающих территорий современной Монголии. Все образцы характеризуются специфичным спектром гаплотипов, подчеркивающим недавний эффект основателя у различных этноспецифичных линий этой гаплогруппы. Оценка возраста по YSNP и YSTR показывает, что представители родов у хакасов, шорцев, ненцев, северных и южных алтайцев являются родственниками по мужской линии и имеют родоначальника, жившего в относительно недалеком прошлом. Эти результаты позволяют гораздо точнее разделить филогенетическую структуру этой гаплогруппы и точнее определить возраст их расхождения, детализировать их происхождение и связь с миграциями их носителей в прошлом.
Финансирование: Исследование поддержано за счет гранта Российского научного фонда (№ 22-64-00060, https://rscf.ru/project/22-64-00060/).

Phylogeny and phylogeography of the Y-chromosome haplog-roup N1a1a1a1a-F1419

Kharkov V.[1]*, Valikhova L.[1], Khitrinskaya I.[1], Zarubin A.[1], Adamov D.[2], Stepanov V.[1]
[1] *Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia*
[2] *M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia*
* *vladimir-kharkov@medgenetics.ru*

Key words: Y-chromosome; haplogroup N1a1; phylogeny; phylogeography

Motivation and Aim: Detailed analysis of the phylogeny of Y-chromosome haplogroups and estimation of the age of different lineages through the evaluation of whole-genome sequencing results contributes significantly to determining their geographical origin, history of demographic events, and migrations. The discovery of novel single nucleotide polymorphisms (SNPs) for various subclades of the Y chromosome allows us to clarify the specific features of all male lineages under study, detailing their ethnic differences and estimating the time of separation of all branches within any haplogroup. Our goal was to conduct a detailed phylogenetic analysis of the North Eurasian N1a1a1a1a-F1419 haplogroup using whole-genome sequences of Y chromosomes from different phylogenetic lineages and isolated samples with specific Y-STR haplotypes. To achieve a more comprehensive phylogenetic analysis of ethnically specific lineages, we increased the number of samples collected from new populations.This allows us to characterize the structure of this haplogroup as accurately as possible by extracting data on new single nucleotide polymorphisms (SNPs) from complete genomes and comparing them with haplotypes using Y-STR markers.
Methods and Algorithms: Haplogroup-specific N1a1a1a1a-F1419 single nucleotide polymorphisms (SNPs) were extracted from genome-wide next-generation sequencing (NGS) data in at least 202 DNA samples from diverse populations in Siberia, Eastern Europe, the Far East, Central Asia, and the North Caucasus. Age estimation was performed using the extracted SNP data. Separate branches were created according to the corresponding samples [1]. The total size of the nucleotide indices calculated for each sample within the combBED [2]. Another tool was created using the hg38

coordinate system, which is part of the combBED. The calculation of the probability of these changes occurring is based on the assumption that Poisson surgery is a mutant [3]. DNA haplotypes were analyzed using the following markers: DYS19, 385a, 385b, 388, 389I, 389II, 390, 391, 392, 393, 426, 434, 435, 436, 437, 438, 439, 442, 444, 445, 448, 449, 456, 458, 460, 462, 481, 504, 505, 518, 525, 531, 533, 537, 552, 570, 576, 635, 643, YCAIIa, YCAIIb, GATA H4.1, Y-GATA-A10, GGAAT1B07. The age estimates were compared with the data from [4] and YFull calculations.

Results: The phylogeny and geographic distribution of specific subclades of the N1a1a1a1a-F1419 haplogroup of the Y chromosome are described by expanding the range of single nucleotide polymorphisms (SNPs). A large number of novel, previously unreported SNPs have been identified in many samples from men belonging to the Siberian and European populations. These SNPs account for approximately 20 % of the total set of markers, but in some samples from ethnic subgroups, their proportion among terminal SNPs can reach 100 %. These findings allow for more accurate identification and differentiation of its variants among men from different genders and ethnic backgrounds. Based on the identification of new YSNPs and the linkage of samples to specific subclusters within various haplogroups, an age estimate was calculated using a standard formula that takes into account the previously reported mutation rate of SNPs, as reported in various articles and on the YFull database.We performed genotyping of a wide range of YSNPs that were informative for detailed analysis of this haplogroup, extracted from sequenced samples using available population samples. Most of the SNPs had their alleles fully confirmed in the already sequenced samples and in all other samples from these haplogroups. It has been shown that most of the newly identified YSNPs in combination with YSTR haplotypes have a high level of ethno-populational specificity. In most cases, a comparative analysis of randomly selected samples allows for an ethno-geographical reference of a sample with accuracy in terms of ethnicity, sub-ethnicity, or clan, and in some cases, even to a specific location and carriers of YSTR haplotypes. A phylogenetic tree of this haplogroup was constructed using 202 samples and median YSTR networks for its various branches from the common ancestor of this haplogroup to individual samples of men with specific terminal YSNPs.As an example, we can estimate the age of the specific Yakut-Evenki lineage, N-M1984, at 960 years old. The age of an earlier lineage, N-M2016, which includes individual samples from Uzbeks, Khakass, and Yakuts, is estimated at 2,520 years old. The total estimated age of haplogroup N1a1a for all sequenced samples from Siberian, Central Asian, and European populations is 14,020 years old. The age of the Buryat lineage is 980 years old, the Siberian Tatar lineage is 1,400 years old, and the Khakass and Shor lineage is 740 years old. For most genotyped samples, a specific link to a specific sublineage has been obtained, which corresponds to their ethnic and subethnic affiliation. As an example of this, we can refer to data on differences in Y-chromosome haplogroups between tundra and forest Nenets populations. Among forest Nenets, the haplogroup N1a1a1a1a2a1c1 (Y13852) is dominant, and all men from the Pyak clan belong to this group. The age of the Pyak clan of Forest Nenets is 451 years, while the tundra Nenets Salinder from Vanuito Phratry is 331 years old and the Tibich clan is 216 years old. These ages agree well with the data provided by anthropologists and ethnographers. The ancestral lineage N-Z35179 among the Komi, Mari, and Udmurt people is 2,800 years old, while the European lineage N-Z35198 common to the Mari and Udmurs is 1,660 years old. The line N-Y40117, which includes three Komi samples, is only 200 years old. Among the three Udmurt samples, the age of N-Z35193 is 300 years.

Overall, the results from calculating the ages of branches of haplogroup N1a1a1a1a-F1419 agree with the data from the compared studies, although there are some differences in the exact ages in some cases. Some of our samples have formed separate sublineages that were not previously described and allow us to better understand the phylogeny of these lineages. A study was conducted to map the main lines of the haplogroup N1a1a1a1a-F1419, based on their frequency in the populations under study. *Conclusion*: The phylogeny of N1a1a1a1a-F1419 haplogroup confirms that its initial distribution in Siberia and Europe occurred in the southern part of Siberia or neighboring regions of modern Mongolia. All the samples share a specific set of haplotypes, indicating a recent founder effect within various ethnically specific lineages of this group. Based on YSNP and YSTR analysis, the age estimates suggest that members of the Khakass, Shors, Nenets, Northern and Southern Altaians share a common ancestor who lived relatively recently in the past. This allows us to better understand the phylogenetic relationships within this haplogroup and determine the time of divergence more precisely. It also helps us to trace their origins and connections with past migrations.

Funding: The study is supported by a grant from the Russian Science Foundation (No. 22-64-00060, https://rscf.ru/project/22-64-00060/).

Список литературы/References

1. Adamov D.S. The Yakut branch of Y-chromosome as a part of the haplogroup N-M2016. *Sib Res*. 2022;2(8):29-36. doi 10.33384/26587270.2022.08.02.05e
2. Adamov D., Guryanov V., Karzhavin S., Tagankin V., Urasin V. Defining a new rate constant for Y-chromosome SNPs based on full sequencing data. *Russ J Genet Genealogy*. 2015;7(1):68-89
3. Poznik D., Henn B., Yee M.-C. et al. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science*. 2013;341(6145):562-565. doi 10.1126/science.1237619
4. Ilumäe A.M., Reidla M., Chukhryaeva M. et al. Human Y chromosome haplogroup N: A non-trivial time-resolved phylogeography that cuts across language families. *Am J Hum Genet*. 2016;99(1):163-173. doi 10.1016/j.ajhg.2016.05.025

doi 10.18699/bgrs2024-4.1-11

Генофонд митохондриальной ДНК и Y-хромосомы населения Южной Сибири гунно-сарматского времени (конец I тыс. до н.э. – первая половина I тыс. н.э.)

Черданцев С.

Институт цитологии и генетики СО РАН, Новосибирск, Россия
stephancherd@gmail.com

Ключевые слова: древняя ДНК; митохондриальная ДНК; Y-хромосома; Южная Сибирь; гунно-сарматское время

Мотивация и цель: Согласно археологическим и историческим данным, кочевые племена конца I тыс. до н.э. – начала I тыс. н.э. играли ключевую роль в культурных и этнических процессах в Центральной Азии и прилегающих районах. Процессы, протекавшие в этот период, определили структуру генофонда многих групп коренного населения Евразии, включая современные монголоязычные и тюркоязычные народы. Исследуемые нами группы населения Южной Сибири железного века (хунну и синхронные им группы населения с территории Тувы и Саяно-Алтайской горной страны) участвовали в масштабных миграционных событиях, охвативших весь евразийский степной пояс и прилегающие к нему регионы. Целью данного исследования является реконструкция процессов формирования генофонда митохондриальной ДНК и Y-хромосомы локально-территориальных кочевых групп Южной Сибири гунно-сарматского времени (хунну Забайкалья, население гунно-сарматского времени Тувы, носители булан-кобинской культуры Горного Алтая, население таштыкской культуры Минусинской котловины).
Методы и алгоритмы: Был исследован генофонд мтДНК (в сумме 205 образцов) и Y-хромосомы четырех групп населения гунно-сарматского времени: хунну с территории Забайкалья, население гунно-сарматского времени Тувы, носителей таштыкской культуры Минусинской котловины и булан-кобинской культуры Горного Алтая.
Результаты: Исследованы структура и филогенетическое положение мтДНК для 18 образцов населения хунну Забайкалья, 35 представителей населения гунно-сарматского времени Тывы, 48 индивидов таштыкской культуры и 104 образца представителей булан-кобинской культуры Горного Алтая. Высокая доля успешно проанализированных образцов (более 70 % от взятых в исследование) свидетельствует о высокой сохранности древней ДНК в антропологическом материале. Данные о вариабельности мтДНК в генофонде всех перечисленных популяций свидетельствуют о высоком генетическом разнообразии популяций. Это подразумевает участие генетически контрастных компонентов в формировании их генетического состава. Получены данные о вариабельности линий Y-хромосомы в мужском генофонде хунну Забайкалья, населения гунно-сарматского времени Тувы, носителей булан-кобинской культуры Горного Алтая и таштыкской культуры Минусинской котловины, позволяющие предварительно

оценить специфику процессов формирования мужской части каждой из анализируемых популяций гунно-сарматского времени.
Выводы: По структуре генофонда мтДНК хунну Забайкалья обнаружили высокое сходство с современными монголоязычными популяциями восточных районов Центральной Азии. Основным механизмом формирования структуры генофонда мтДНК других (помимо хунну) рассмотренных в данном исследовании популяций Южной Сибири гунно-сарматского времени является их генетическая преемственность с предшествовавшим населением скифского времени. Таким образом, наблюдаемый уровень генетического разнообразия населения в значительной степени сформировался еще в предшествующее скифское время. Наряду с этим, в генофонде мтДНК исследованных групп населения Южной Сибири присутствуют минорные компоненты, которые указывают на потенциальное генетическое влияние хунну и/или родственных им групп кочевников Центральной Азии. Это влияние, по-видимому, играло роль второстепенного фактора при формировании генофонда мтДНК населения Алтае-Саянской горной системы и прилегающих регионов. Полученные данные по мужскому генофонду в целом подтверждают эти выводы.
Финансирование: Исследование поддержано бюджетным проектом ИЦиГ СО РАН (№ FWNR-2022-0011).

Mitochondrial DNA and Y-chromosome gene pool of Xiongnu-Sarmatian period populations from South Siberia (the end of I millennium B.C. until I millennium A.D.)

Cherdantsev S.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
stephancherd@gmail.com

Key words: ancient DNA, mitochondrial DNA, Y-chromosome, South Siberia, Xiongnu-Sarmatian period

Motivation and Aim: According to archaeological and historical data, nomad tribes of I millennium B.C. until I millennium A.D. played a key role in cultural and ethnic processes in Central Asia and adjacent regions. The processes that took place during this period determined the gene pool structure of the Eurasia indigenous population, including modern Mongol- and Turkic-speaking ethnic groups. The Southern Siberia Iron Age population under study (the Xiongnu and synchronous groups of Tuva and the Sayan-Altai mountainous country) participated in large-scale migration events that covered the entire Eurasian steppe belt and adjacent regions. The aim of this study is the reconstruction of processes, which shaped the mitochondrial DNA and Y-chromosome gene pool of local nomad populations dated by Xiongnu-Sarmatian period in the regions of Mountain Altai, Minusinsk Basin and Tuva based on the mtDNA and Y-chromosome haplogroups variation data.
Methods and Algorithms: MtDNA (205 samples) and Y-chromosome gene pool of four Xiongu-Sarmatian time populations' were studied – the Xiongnu from the territory of Transbaikalia, the Tuva population of the Xiongnu-Sarmatian time, the carriers of the Minusinsk Basin Tashtyk culture and the Altai Mountains Bulan-Koba culture.

Results: The mtDNA structure and phylogenetic position were studied for the Transbaikalian Xiongnu population (18 samples), Xiongnu-Sarmatian time of Tuva population (35), Minusinsk Basin Tashtyk culture population (48) and the Altai Mountains Bulan-Koba culture population (104). The high proportion of successfully analyzed samples (more than 70 % of those taken for the study) indicates the high preservation of ancient DNA in the anthropological material. MtDNA variability data in the gene pool of all populations under study indicate high genetic diversity. This implies the genetically contrasting components participation in their genetic composition formation. The all populations under study Y-chromosome variability data were obtained. This allowing a preliminary assessment of the male part gene pool formation processes specificity each of Xiongnu-Sarmatian period populations under study.
Conclusion: According to the mtDNA gene pool structure, the Transbaikalian Xiongnu found high similarity with modern Mongol-speaking populations of Central Asia eastern regions. The main mechanism for the mtDNA structure gene pool formation of other Southern Siberia Xiongnu-Sarmatian time populations (besides the Xiongnu) considered in this formation is their genetic continuity with the previous Scythian time population. Thus, the observed level of genetic diversity in a population was largely formed in the previous Scythian time. The mtDNA gene pool of the experimental Southern Siberia population groups contains minor components that indicate the probable genetic influence of the Xiongnu and/or related Central Asian nomadic groups. This influence apparently played the role as a secondary factor in changing the MtDNA gene pool of the Altai-Sayan mountain system and nearby regions population. The data obtained on the male gene pool generally confirm these conclusions.
Funding: The study is supported by government IC&G SB RAS project (No. FWNR-2022-0011).

doi 10.18699/bgrs2024-4.1-12

Application of biobanking for personal identification in the Russian Federation

Faleeva T.G.

North-Western State Medical University named after I.I. Mechnikov, St. Petersburg, Russia
St. Petersburg State Budgetary Healthcare Institution "Bureau of Forensic Medicine", St. Petersburg, Russia
Tatiana.fal@mail.ru

Key words: biobank; biobanking; DNA personal identification; genetic certification of the population; proactive DNA registration

Motivation and Aim: Every year in Russia the number of disappearances of people increases. According to official data, the number of missing persons in Russia in 2023 amounted to tens of thousands of people. However, most cases of unknown disappearance remain unsolved. The peculiarities of the geopolitical situation in the country and the world determine the need for the formation of proactive DNA certification of the population in order to increase the efficiency of the personal identification process – one of the main tasks of forensic medicine. The organization of a large-scale system for collecting, storing, processing and using biological material in research based on biobanks will allow not only to obtain new medical knowledge about the state of human health for scientific and therapeutic purposes, but also to use such data in forensic medical expert practice. The purpose of this study is to analyze the current state of biobanking in our country and the possibility of using its resources for DNA identification in the structure of the forensic medical service.
Methods and Algorithms: Official data regarding missing citizens was analyzed. The regulatory framework for the system of organizing and managing the processes of collecting and using genetic information in Russia has been studied. The features of the functioning of biobanks in the country and the world were studied, and the possibilities of their use within the framework of personal identification were assessed.
Results: In the field of forensic medicine, various methods of personal identification are used, among which molecular genetic research stands out as a particularly evidentiary method. Direct DNA identification, considered the most reliable technique, allows genetic material of unknown origin to be compared with a person's already known genetic profile. This process is made possible by proactive DNA recording, which involves the collection, storage and analysis of comparative biological samples. The importance of genetic registration is especially high for people whose professional activities involve an increased risk to life.
The development of genomic registration in Russia is actively supported by the Government, as reflected in Federal Law No. 242-FZ of December 3, 2008 [1], which details the principles and procedures of genomic registration, processing of biological material and genetic information, as well as control measures and supervision of the implementation of these processes. The information collected under this law is intended to be used for the purpose of preventing, detecting and investigating crimes, searching for missing people, both citizens of the Russian Federation and foreigners located in the country, identifying unidentified persons and corpses, as well as determining family ties.

State genomic registration is mandatory for certain categories of the population: convicts, persons prosecuted, as well as unidentified bodies. For other citizens, genomic registration is voluntary and is carried out for a fee, which hinders the widespread implementation of genomic certification. There are also restrictions on the period of storage of genetic information: for unidentified persons and unidentified corpses, the period is 70 years, for others – until the fact of death is established or until the age of 100 is reached in the absence of information about death. The possibility of destruction of genomic information is provided upon application by a registered person and on the basis of a court order.

Decree of the President of Russia dated March 11, 2019 No. 97 defines strategic directions for ensuring chemical and biological safety in the country [2]. The document clarifies that restrictions on personal rights and freedoms in the interests of chemical and biological safety are permitted only to the extent necessary to protect public health, their rights and interests, as well as to guarantee the defense capability and security of the state. The fundamental objectives include genetic certification of citizens in compliance with legal guarantees for the protection of information about the human genome and the creation of genetic profiles of the population. The importance of improving legislation and management mechanisms in this area is also emphasized. The key tasks within the framework of ensuring the resource base of the national chemical and biological safety system are the creation of conditions for genetic certification of the population and the development of screening technologies for the study of gene pools of humans, animals and plants.

In situations where direct DNA identification is not possible due to the lack of samples for comparison, the method of indirect DNA identification is successfully used. This method involves analyzing and comparing the genetic profiles of close relatives of a person reported missing with the genetic profile of an unidentified individual. Thus, Federal Law No. 16-FZ of February 14, 2024 [3] amends the previously existing Federal Law "On State Genomic Registration in the Russian Federation" [1]: Article 7 added a clause according to which the mandatory state genomic registration Now close relatives of persons officially recognized as missing are also subject to this.

The implementation of planned initiatives and projects requires coordinated and balanced approaches to solving a set of related and dependent tasks in various areas, including in the aspect of personal identification in forensic practice. At the same time, each of the key areas requires careful analysis and implementation of models, mechanisms, tools and methods used. One of these specific but important issues is the development of biobanking in Russia.

Human biological specimen biobanks are organizations or institutions that collect, store, and make available biological specimens such as blood, saliva, DNA, tissue, cells, etc., for research purposes. They aim to accelerate progress in scientific research and ensure high quality samples for the scientific community. There are many such biobanks around the world, with some specializing in the study of specific diseases or populations, while others support a diversity of research in a wide range of areas.

Although the creation of a specialized system for collecting, storing, processing, analyzing and utilizing genetic information in the form of a state biobank seems logical, in the Russian Federation there are obstacles that hinder its development. Challenges include the lack of clear regulations governing biobanking compared to classical institutions. According to experts, current technical standards for storing biological samples are insufficient.

Moreover, the Russian Federation has introduced GOST R ISO 20387-2021 [4], which is a national standard in the field of biotechnology and biobanking. This standard defines key terminology, establishes the basic legal and organizational principles of biobank activities, requirements for personnel, equipment, procedures for collecting and storing samples, as well as for quality control and record keeping. Despite its wide coverage, GOST does not cover standards for the disposal of biomaterials, with the exception of rules for maintaining documentation related to disposal.

The next significant obstacle to the development of biobanking is the ethical issues associated with the use of biological material and medical information of citizens for research purposes, even under conditions of anonymity. The key aspect here is the development and implementation of legal regulations regarding the processing of anonymized data, ensuring their confidentiality, protection and lawful use of human tissue in scientific work, which is a complex task.

Conclusion: The statistics of missing persons in Russia in recent years makes us think about the existing problems in society and the mechanisms for its protection. The formation of a unified state system for the collection, storage and research of human biological material, including DNA identification of an individual, is the highest priority task in our country, for the implementation of which there is sufficient potential and a scientific base. An analysis of the existing biobanking system has shown that biobanks are a key tool for the progress of modern medicine in general, including the creation of medicines and diagnostic methods, and represent the most important resource for personalized medicine and forensic identification examination. However, unlike a number of European countries, Russia does not have special laws on biobanking. Existing legal norms are "built-in" into existing fundamental regulations on public health and information protection. This state of affairs reveals the need to create and modernize the regulatory framework governing the functioning of biobanks in the Russian Federation. The development of state genomic registration, ensuring the establishment of proactive DNA certification and the creation of specialized institutions – biobanks – will increase the efficiency and speed of personal identification, the detection of crimes, the objectivity of the examinations performed, and will also ensure the safety and improve the quality of life of all citizens.

Funding: The study is supported by self-funding.

References

1. ICRC Database, National Practice, Russian Federation, Federal Law No. 242 FZ on state genome registration, 2008. https://ihl-databases.icrc.org/en/national-practice/federal-law-no-242-fz-state-genome-registration-2008 (accessed on 31.05.2024)
2. Decree of the President of the Russian Federation of March 11, 2019 No. 97 On the Fundamentals of State Policy of the Russian Federation in the field of ensuring chemical and biological safety for the period until 2025 and beyond. *Law Enforcement Rev.* 2023;7(2):105-111 (in Russian)
3. Federal Law of February 14, 2024 No. 16-FZ On Amendments to the Federal Law On State Genomic Registration in the Russian Federation (in Russian). http://publication.pravo.gov.ru/document/0001202402140006
4. GOST R ISO 20387-2021 National standard of the Russian Federation. Biotechnology. Biobanking. General requirements (in Russian). https://docs.cntd.ru/document/1200181384

doi 10.18699/bgrs2024-4.1-13

HSPA8 gene polymorphisms and the risk of severe COVID-19

Karpenko A.[1, 2], Kobzeva K.[2], Sergeeva V.[1], Bushueva O.[2, 3]*

[1] *Department of Anesthesia and Critical Care, Institute of Continuing Education, Kursk State Medical University, Kursk, Russia*

[2] *Laboratory of Genomic Research, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russia*

[3] *Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, Kursk, Russia*

* *olga.bushueva@inbox.ru*

Key words: heat shock; severe COVID-19; rs1461496; rs1136141

Motivation and Aim: COVID-19, caused by the novel coronavirus SARS-CoV-2, has varying degrees of severity, from mild or asymptomatic cases to critical symptoms such as respiratory failure and multiple organ dysfunction. Heat shock proteins (HSPs) are induced by cellular stressors such as inflammation and infection and are involved in cellular homeostasis and immune responses. HSPs have the ability to protect cells and tissues from the harmful effects of inflammation, and HSP70 may have a protective effect on the immune system by promoting antigen processing and presentation. HSPA8, a member of the heat shock protein 70 (HSP70) family, has been implicated in various stages of the viral life cycle, including attachment [1, 2], internalization [3], and replication [4]. The aim of our study was to evaluate the impact of polymorphic variants rs1461496 and rs1136141 *HSPA8*, member of the HSP70 family, on the severity of COVID-19 in the population of Central Russia.

Methods and Algorithms: The study included 1,152 unrelated individuals from Central Russia, including 199 patients with COVID-19 hospitalized in the intensive care units of Kursk hospitals, and 905 patients in the control group with a mild form of COVID-19 that did not require hospitalization. Genotyping of *HSPA8* SNPs was carried out using real-time PCR. A log-additive regression model was used to assess associations. Functional annotation of SNPs was performed using a range of bioinformatics resources.

Results: Our study identified that the rs1461496 *HSPA8* increased the risk of severe COVID-19 in patients aged 68 and older (risk allele A, OR = 1.61, 95 % CI 1.06–2.44, p = 0.024), while also decreasing the time to the onset of clot growth in the overall group (Tlag, minutes, P = 0.02). This SNP regulates the expression via cis-eQTL effects of *CRTAM*, that was found to be down-regulated in COVID-19 patients. *CRTAM* plays a role in the activation and differentiation of several T-cell subsets, including NK cells [5]. Moreover, analysis of transcription factors (TF) binding to SNP, revealed that the risk allele A rs1461496 *HSPA8* is implicated in various pathological pathways crucial for COVID-19, such as positive regulation of viral transcription by the host, lymphocyte differentiation, and the canonical Wnt signaling pathway, which can lead to a cytokine storm [6]. In contrast, the protective allele G rs1461496 *HSPA8* creates DNA binding sites for TFs involved in biological processes that may positively impact COVID-19 outcomes. These processes include the positive regulation of nitric oxide (NO) biosynthesis, regulation of transforming growth factor beta (TGF-β) signaling and production, and response to hypoxia. Firstly, NO operates in COVID-19 through four mechanisms: regulating blood flow, initiating anti-inflammatory responses, promoting

anti-coagulation effects, and exerting antiviral properties [7], thus indicating the protective effect of this SNP. Secondly, TGF-β influences immune cell development, differentiation, tolerance induction, and homeostasis. A study found that serum levels of TGF-β positively correlated with improved outcomes in COVID-19 patients [8].

Additionally, we identified that rs1136141 *HSPA8* increased the risk of a severe COVID-19 course in patients under 68 years old (risk allele A, OR = 1.55, 95 % CI 1.06–2.28, p = 0.03). Data from the Lung Knowledge Portal also associate rs1136141 *HSPA8* with reduced forced expired volume in 1 second (FEV1), forced vital capacity (FVC), and peak expiratory flow. Furthermore, the risk allele A rs1136141 *HSPA8*, through TF binding, is involved in leukocyte differentiation and hippo signaling, inhibition of which significantly reduces SARS-CoV-2 replication [9]. Conversely, the protective allele G rs1136141 *HSPA8* positively regulates CD8-positive, alpha-beta T cell differentiation, negatively regulates leukocyte cell-cell adhesion, and is involved in response to cAMP, which can prevent antibody-mediated coagulopathy in COVID-19 [10]. Moreover, rs1136141 *HSPA8* (p = 0.007) and rs10892958 *HSPA8* (p = 0.007) were associated with lower ground-glass opacity upon admission to the ICU, and these same SNPs, rs1136141 *HSPA8* (p = 0.025) and rs10892958 *HSPA8* (p = 0.015), showed associations with reduced ground-glass opacity upon discharge from the ICU.

Conclusion: Thus, our study represents the first investigation into the associations of *HSPA8* SNPs rs1461496 and rs1136141 with severe COVID-19 within the Caucasian population of Central Russia. We conducted a comprehensive analysis of the impact of *HSPA8* loci on the clinical manifestation of the disease, thrombodynamic parameters, and the modifying effects of age on the risk of severe COVID-19, providing novel insights into the genetic factors underlying disease severity in this population.

Funding: The study is supported by Kursk State Medical University.

References

1. Watanabe K., Fuse T., Asano I. et al. Identification of Hsc70 as an influenza virus matrix protein (M1) binding factor involved in the virus life cycle. *FEBS Lett.* 2006;580(24):5785-5790. doi 10.1016/j.febslet.2006.09.040
2. Zhu P., Lv C., Fang C. et al. Heat shock protein member 8 is an attachment factor for infectious bronchitis virus. *Front Microbiol.* 2020;11:1630. doi 10.3389/fmicb.2020.01630
3. Zárate S., Cuadras M.A., Espinosa R. et al. Interaction of rotaviruses with Hsc70 during cell entry is mediated by VP5. *J Virol.* 2003;77(13):7254-7260. doi 10.1128/jvi.77.13.7254-7260.2003
4. Salinas E., Byrum S.D., Moreland L.E., Mackintosh S.G., Tackett A.J., Forrest J.C. Identification of viral and host proteins that interact with murine gammaherpesvirus 68 latency-associated nuclear antigen during lytic replication: a role for Hsc70 in viral replication. *J Virol.* 2016;90(3):1397-1413. doi 10.1128/JVI.02022-15
5. Alqutami F., Senok A., Hachim M. COVID-19 Transcriptomic atlas: a comprehensive analysis of COVID-19 related transcriptomics datasets. *Front Genet.* 2021;12:755222. doi 10.3389/fgene.2021.755222
6. Vallée A., Lecarpentier Y., Vallée J.-N. Interplay of opposing effects of the WNT/β-catenin pathway and PPARγ and implications for SARS-CoV2 treatment. *Front Immunol.* 2021;12:666693. doi 10.3389/fimmu.2021.666693
7. Rajendran R., Chathambath A., Al-Sehemi A.G. et al. Critical role of nitric oxide in impeding COVID-19 transmission and prevention: a promising possibility. *Environ Sci Pollut Res.* 2022;29(26):38657-38672. doi 10.1007/s11356-022-19148-4
8. Zivancevic-Simonovic S., Minic R., Cupurdija V. et al. Transforming growth factor beta 1 (TGF-β1) in COVID-19 patients: relation to platelets and association with the disease outcome. *Mol Cell Biochem.* 2023;478(11):2461-2471. doi 10.1007/s11010-023-04674-7
9. Jr G.G., Jeyachandran A.V., Wang Y. et al. Hippo signaling pathway activation during SARS-CoV-2 infection contributes to host antiviral response. *PLoS Biol.* 2022;20(11):e3001851. doi 10.1371/journal.pbio.3001851
10. Zlamal J., Althaus K., Jaffal H. et al. Upregulation of cAMP prevents antibody-mediated thrombus formation in COVID-19. *Blood Adv.* 2022;6(1):248-58. doi 10.1182/bloodadvances.2021005210

doi 10.18699/bgrs2024-4.1-16

GWAS-significant loci and risk of uterine fibroids in the population of Central Russia

Ponomareva L.[1], Babkina M.[2], Kobzeva K.[2], Bushueva O.[2, 3]*

[1] *Department of Obstetrics and Gynecology, Institute of Continuing Education, Kursk State Medical University, Kursk, Russia*

[2] *Laboratory of Genomic Research, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russia*

[3] *Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, Kursk, Russia*

* *olga.bushueva@inbox.ru*

Key words: SNP; uterine fibroid; GWAS

Motivation and Aim: Uterine fibroids (UF) are common gynecological issue that affects up to 70 % of women of reproductive age [1, 2]. Surgery is currently the main treatment procedure [3]. At the moment, the age of primary diagnosis of uterine fibroids is becoming younger [1, 3]. The disease affects their ability to conceive and have successful pregnancies. Since this condition has a permanent impact on a woman's ability to have children, it's crucial to pay close attention to the factors that increase the risk of UF development. A lot of identified genetic variations associated with UF were identified by GWAS [4, 5]. The aim of this study was to explore the association between GWAS-significant SNPs and the susceptibility to UF in the population of Central Russia.
Methods and Algorithms: DNA samples from 654 UF patients and 401 healthy individuals were analyzed using real-time PCR to detect the rs547025 *SIRT3*, rs2456181 *ZNF346*, rs7907606 *SLK, STN1*, rs58415480 *SYNE1*, rs7986407 *FOXO1*, rs72709458 *TERT*, rs117245733 *LINC00598* SNPs, was identified in previous GWASs as UF-linked loci. A log additive regression model was used to assess the associations. Additionally, bioinformatics tools were utilized to investigate the functional effects of the SNPs.
Results: We observed a significant link between the rs547025 *SIRT3* polymorphism and a decreased risk of UF development (OR = 0.64, 95 % CI = 0.47–0.88, P = 0.005). Further investigation using HaploReg v4.2 revealed that rs547025 *SIRT3* resides in the DNA binding region of H3K4me1, with its impact strengthened by H3K27ac, indicating its significant epigenetic regulatory potential in blood. Moreover, rs547025 displays notable cis-eQTL effects, influencing the expression of several genes, including *RIC8A*, *PSMD13*, *BET1L*, *SCGB1C1*, *ODF3*, *IFITM2*, *IFITM1*, *IFITM3*, and *PTDSS2* in blood and adipose tissue. Interestingly, *BET1L* has previously been linked to UF risk, while *IFITM1* is a sensitive marker for endometriosis. These findings offer valuable insights into potential pathways influenced by the *SIRT3* polymorphism in UF development. Additionally, data from the Reproductive System Knowledge Portal indicates a significant association between the rs547025 variant and earlier age at natural menopause, which is recognized as a significant risk factor for UF.
Conclusion: To sum it up, this study was the first to establish the significant association of rs547025 *SIRT3* with the risk of UF in the population of Central Russia. Therefore, this locus can be used as an effective predictive marker of UF development at the preclinical stage as part of predictive medicine.
Funding: The study is financially supported by Kursk State Medical University.

References

1. Lou Z., Huang Y., Li S. et al. Global, regional, and national time trends in incidence, prevalence, years lived with disability for uterine fibroids, 1990–2019: an age-period-cohort analysis for the global burden of disease 2019 study. *BMC Public Health*. 2023;23:916. doi 10.1186/s12889-023-15765-x
2. Sliz E., Tyrmi J.S., Rahmioglu N. et al. Evidence of a causal effect of genetic tendency to gain muscle mass on uterine leiomyomata. *Nat Commun*. 2023;14(1):28-35. doi 10.1038/s41467-023-35974-7
3. Karamyan R.A., Ordiyats I.M., Khorolskiy V.A., Asatryan D.R. Uterine fibroids: a look at the problem. *Med Her South Russia*. 2022;13(2):18-25. doi 10.21886/2219-8075-2022-13-2-18-25 (in Russian)
4. Dai Y., Liu X., Zhu Y. et al. Exploring Potential Causal Genes for Uterine Leiomyomas: A Summary Data-Based Mendelian Randomization and FUMA Analysis. *Front Genet*. 2022;13:890007. doi 10.3389/fgene.2022.890007
5. Gallagher C.S., Mäkinen N., Harris H.R. et al. Genome-wide association and epidemiological analyses reveal common genetic origins between uterine leiomyomata and endometriosis. *Nat Commun*. 2019;10(1):48-57. doi 10.1038/s41467-019-12536-4

doi 10.18699/bgrs2024-4.1-17

Phylogenetic analysis of the tumor-specifically expressed pseudogene ENSG00000186076

Potter I.[1]*, Zykova M.[1], Makashov A.[2]

[1] *State budgetary educational institution secondary school No. 225, St. Petersburg, Russia*

[2] *Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

** vanekpotter@gmail.com*

Key words: tumor; expressions; pseudogene; non-coding; phylogenetic analysis

Motivation and Aim: In this paper, we performed a phylogenetic analysis of the non-coding gene ENSG00000186076. This gene is located at chromosome 12 and has been identified as a processed pseudogene of PGAM1 gene. According to Bibert et al, 2021 the ENSG00000186076 gene expression is significantly reduced in a group of patients with severe COVID-19, indicating the level of expression of this gene is associated with the severity of disease. The homologs were searched in 15 genomes of species of the human lineage.
Methods and Algorithms: To proceed analysis and build phylogenetic tree were used full genomes of 15 species, sequence of non-coding gene. We then performed phylogenetic tree reconstruction to examine the evolutionary relationships among the selected sequences using multiple alignment and based on the alignment tree was constructed. For tumor expression analysis, expressions of this gene in different tumors were used.
Results: The phylogenetic tree showed that this gene has a lot of copies or that it's a copy of another. Expression analysis showed reliable differences in tumor expression.
Conclusion: This gene followed our lineage for a long time and presumably has impact on tumor forming.
Funding: The study is not supported financially.

References

1. Кребс Д., Голдшейн Э., Килпатрик С. Гены по Льюину. М.: Лаборатория знаний, 2020
2. Bibert S., Guex N., Lourenco J. et al. Transcriptomic Signature Differences Between SARS-CoV-2 and Influenza Virus Infected Patients. *Front Immunol*. 2021;12:666163. doi 10.3389/fimmu.2021.666163

doi 10.18699/bgrs2024-4.1-18

On to the issue of the oldest population of North Asia

Shunkov M.
Institute of Archaeology and Ethnography, SB RAS, Novosibirsk, Russia
shunkov77@gmail.com

Key words: North Asia; Altai; *Homo erectus*; Denisovans; Altai Neanderthals

According to modern data from archaeology, paleoanthropology and paleogenetics, *Homo erectus* left Africa about 1.9 million years ago and began to settle in Eurasia. The first wave of ancient migrants moved in two main directions: the first is through the Middle East to the south of Europe – to the Caucasus and the Mediterranean regions, the second – is through the western regions of Asia to the east. It is assumed that the early hominins moved east in two ways. One of them probably ran south of the Himalayas and Tibet through Hindustan to East and Southeast Asia. Another, northern migration route probably passed through the Central Asian Highlands to Central and North Asia.

The oldest evidence of the appearance of man in the territory of Northern Asia is the materials of the Early Paleolithic site of Karama, located in the north-west of Altai [1]. At Karama, in the sediments of the first half of the Middle Pleistocene, a cultural sequence of four horizons of the Early Paleolithic pebble-chip industry was recorded, the age of which was determined in the chronological range of 800–600 ka BP. Judging by the materials of paleobotanical studies, the process of settlement of Early Paleolithic people in Altai took place in fairly favorable climatic conditions. At that time, the surrounding landscapes were characterized by pine and birch forests with the participation of dark coniferous species, as well as species exotic to the modern Altai flora – elm, hornbeam, hop, linden, maple, oak, Manchurian walnut and hazel [2].

Currently, it can be said that the initial settlement of the territory of North Asia occurred, most likely, from the southwest, or rather from the low-mountain and foothill regions of the Altai, where, in addition to Karama, two other Early Paleolithic sites are known – Ulalinka and MK 1. The archaic pebble industry of Karama indicates the appearance of early hominins here, who were most likely at the stage of *Homo erectus*. Erectus came to the territory of the Altai with the oldest migration wave from Africa, moving in a northeasterly direction through the Middle East, the Central Asian Highlands and the western regions of Central Asia.

Karama's chronostratigraphic column indicates that representatives of the oldest migration stream lived in southern Siberia for almost the entire first half of the Middle Pleistocene, with the exception of the relative cooling phase corresponding to the Mansi horizon (MIS 18). After the cold maximum of the next glacial (MIS 16), due to the general deterioration of the natural environment, early hominins most likely moved to areas with a more temperate climate. The remaining part of the population, apparently, could not adapt to the changed landscape and climatic conditions and ceased to exist.

The next documented period of the ancient history of North Asia is associated with the appearance of Denisovans on the territory of the Altai in the middle of the Middle Pleistocene – carriers of technical and typological traditions based on the Levallois and parallel principles of stone splitting [3]. These processes are the result of the migration of part of the late *Homo heidelbergensis* about 400–350 ka BP from the Levant in an

easterly direction through South and Central Asia, and then to the south of Siberia. This was the second stage in the formation of the modern human species – the division of one ancestral form of the late *H. heidelbergensis* into Neanderthals and Denisovans. The population of late *H. heidelbergensis* settling in the second half of the Middle Pleistocene to the east, came into contact with the descendants of Asian erectus and gave rise to a new taxon – Denisovans, who appeared in the Altai, in the Denisova Cave about 300 ka BP. The fact that the Denisovans are associated with the migration of the late *H. heidelbergensis* to the east can also be evidenced by the earliest industry of the Denisova Cave, which has common features with the Ashelo-Yabroudien materials of the Levant [4]. The Denisovans, with the periodic participation of Altai Neanderthals, developed an original stone industry during the Middle Paleolithic, which served as the basis for the autochthonous formation of the Upper Paleolithic culture about 50 ka BP.

If the formation of Middle Paleolithic cultural traditions begins in the Altai from the middle of the Middle Pleistocene, then in other regions of North Asia this period, up to and including the first cooling of the Upper Pleistocene, is associated with the spread of pebble industries of Early Paleolithic appearance. Such specificity in cultural manifestations is, apparently, a regional reflection of the general patterns of development of the Early Paleolithic Ecumene in the eastern part of Eurasia. The appearance of the pebble industries in the eastern regions of Siberia may have been influenced by Early Paleolithic traditions that developed in the east and southeast Asia.

Funding: The study is supported of a grant from the Russian Science Foundation (No. 24-18-00069).

References

1. Derevyanko A.P., Shun'kov M.V. The Early Palaeolithic Karama Camp in the Altai: The First Results of Studies. *Archaeol Ethnol Anthropol Eurasia*. 2005;3:52-69
2. Bolikhovskaya N.S., Derevyanko A.P., Shunkov M.V. The Fossil Palynological Flora, Geological Age, and Climatic Stratigraphy of the Earliest Deposits of the Karama Site (Early Paleolithic, Altai Mountains). *Paleontol J.* 2006;40:558-566
3. Derevianko A.P., Shunkov M.V., Kozlikin M.B. Who Were the Denisovans? *Archaeology, Ethnology and Anthropology of Eurasia*. 2020;48(3):3-32. doi 10.17746/1563-0110.2020.48.3.003-032
4. Shunkov M.V., Kozlikin M.B. The Earliest Paleolithic Assemblages from Denisova Cave in the Altai. *Archaeology, Ethnology and Anthropology of Eurasia*. 2023;51(1):18-32

doi 10.18699/bgrs2024-4.1-19

Large-scale analysis of haplotype effects in TCR alpha and beta germlines based on Rep-seq data

Vinogradova S.[1#], Bagrova O.[2#], Vlasova E.[3, 4], Shugay M.[3]*

[1] *Moscow Center for Advanced Studies, Moscow, Russia*

[2] *Lomonosov Moscow State University, Moscow, Russia*

[3] *Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russia*

[4] *ITMO University, St. Petersburg, Russia*

[#] *These authors contributed equally to this work*

* *mikhail.shugay@gmail.com*

Key words: T cell repertoires; rare haplotypes; TRA and TRB deletions; V and J gene usage distribution

Motivation and Aim: T cells are an integral part of the human adaptive immune system, with their specificity derived from the unique T cell receptors (TCRs). TCR formation involves a complex process called V(D)J recombination. Through this mechanism, specific V, D, and J genes are chosen from the TRA (TCR alpha) and TRB (TCR beta) loci and are then rearranged to form the TCRα and TCRβ chains of the receptor, respectively [1]. The resulting TCR diversity enables the recognition of a wide array of antigens, facilitating an effective adaptive immune response. To analyse this diversity and track immune responses, a technique called Repertoire sequencing (Rep-seq) is utilised, which involves sequencing of both TCRs and BCRs (B cell receptors) [2]. Recent studies have revealed the presence of both amplifications and deletions within the TRA and TRB genes, as well as in the IGHV loci of BCRs [3, 4]. However, the coherence of these genomic alterations and their biological implications remain unexplored. In this project, we aimed to systematically analyse the TCR repertoires of two large cohorts from Russian [5] and American [6] populations. Our goal was to identify common and rare haplotypes by analysing the frequency of usage of specific V and J genes within individual T cell repertoires. This research not only provides valuable insights into the diversity of the TCR repertoire across these populations but also seeks to explain the observed genomic differences by considering factors such as ethnicity variations across the population, MHC context or thymic selection.

Methods and Algorithms: The study involved two cohorts of donors: Russian donors, consisting of 237 samples subjected to 5'RACE-sequencing, and American donors, with 786 samples obtained from Adaptive Biotechnologies [6]. Analysis was conducted separately within each cohort. Batch effect correction was applied within each group. We normalised gene usage (fraction of reads mapping to a given gene in a given sample) across samples, first standardising to a normal distribution and then scaling back to a [0, 1] interval. Adjusted clonotype frequencies were computed based on these scaled values. We further constructed V/J usage matrices indicating the usage of each segment in samples [5].

Results: We began by analysing the distribution of gene usage across the TCRα and TCRβ chains. While the majority of genes displayed expected unimodal distributions, some revealed bimodal or trimodal patterns. We suggest the bi- or tri- modal distributions are the sign of haplotype effects in populations, appearing due to gene

deletion or homo-/hetero- zygote gene condition. Investigating whether distribution modality could be explained by donor's race, we found statistically significant differences in gene usage between Asian and Caucasian populations for TRBV7-4 gene, and between Latino and Caucasian populations for TRBV4-3 gene. Hierarchical clustering identified rare haplotypes within each population, e.g. TRAJ21, with zero usage in most donors but high usage in a few individuals. Additionally, hierarchical clustering among TRBJ genes revealed two distinct subgroups (J1 and J2) based on usage patterns, possibly influenced by sequence similarity. We further differentiated between functional and non-functional CDR3 sequences, identifying non-functional sequences by stop codons or frame shifts within the CDR3 region. We observed that certain J genes showed a higher usage of non-functional sequences compared to functional ones. While some of these genes are recognised pseudogenes, others without known non-functionality displayed similar patterns, suggesting negative thymic selection against such CDR3 configurations.

Conclusion: This study systematically analysed TCR repertoires in large cohorts from Russian and American populations, revealing significant variations in V and J gene usage across TCRα and TCRβ chains, potentially influenced by ethnicity. Hierarchical clustering uncovered rare haplotypes, indicating unique immune repertoire characteristics within subgroups of donors. Additionally, our investigation into functional versus non-functional CDR3 sequences suggested certain V genes associated with higher frequency non-functional sequences may undergo negative thymic selection, implying a possible biological mechanism where certain gene configurations are selectively disadvantageous. Such T cell repertoire analysis is crucial for practical applications, as understanding the nuances of adaptive immunity can directly influence the selection of targets for receptor-based therapies.

Funding: This work was supported by a grant from the Ministry of Science and Higher Education of the Russian Federation (075-15-2019-1789).

References

1. Murphy K., Weaver C. Janeway's Immunobiology. Garland Science, 2016. doi 10.1201/9781315533247
2. Benichou J., Ben-Hamo R., Louzoun Y., Efroni S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology.* 2012;135(3):183-191. doi 10.1111/j.1365-2567.2011.03527.x
3. Omer A., Peres A., Rodriguez O.L. et al. T cell receptor beta germline variability is revealed by inference from repertoire data. *Genome Med.* 2022;14(1):2. doi 10.1186/s13073-021-01008-4
4. Kidd M.J., Chen Z., Wang Y. et al. The Inference of Phased Haplotypes for the Immunoglobulin H Chain V Region Gene Loci by Analysis of VDJ Gene Rearrangements. *J Immunol.* 2012;188(3):1333-1340. doi 10.4049/jimmunol.1102097
5. Vlasova E.K., Nekrasova A.I., Komkov A.Y. et al. Robust detection of SARS-CoV-2 exposure in the population using T-cell repertoire profiling. *BioRxiv.* 2023. doi 10.1101/2023.11.08.566227
6. Emerson R.O., DeWitt W.S., Vignali M. et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet.* 2017;49(5):659-665. doi 10.1038/ng.3822

4 Симпозиум «Эволюционная, популяционная и медицинская геномика/генетика человека: компьютерные и экспериментальные подходы»

Symposium “Evolutionary, population and medical genomics/genetics of human: computational and experimental approaches”

doi 10.18699/bgrs2024-4.2-01

Обобщение внутрилабораторных критериев качества геномных вариантов, валидируемых по результатам полногеномного секвенирования

Антоненко А.[1]*, Белов Р.[1], Ревкова М.[1], Доморацкая Е.[1], Барциц А.[1], Панферова А.[1], Губона М.[1], Уланова П.[1], Дибирова Х.[1], Соколова Н.[1], Дорощук Н.[1], Макарова М.[1], Грознова О.[2], Криницына А.[1], Беленикин М.[1]**

[1] *ООО «Эвоген», Москва, Россия*

[2] *Научно-исследовательский клинический институт педиатрии им. акад. Ю.Е. Вельтищева ФГБОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России, Москва, Россия*

** antonenko_a.n@mail.ru; ** evogenlab@yandex.ru*

Ключевые слова: полногеномное секвенирование; WGS; валидация вариантов NGS; секвенирование по Сэнгеру

Мотивация и цель: Стремительное развитие технологий высокопроизводительного секвенирования (NGS), алгоритмов биоинформатики и обработки данных NGS сопряжено с необходимостью валидации выявленных потенциальных причинных (каузативных) вариантов альтернативными методами. Наиболее часто валидация найденных вариантов проводится с помощью капиллярного секвенирования (КС) методом Сэнгера, так как данный метод считается так называемым золотым стандартом при проведении клинических исследований. Количество научных статей по сравнению результатов NGS для экзомов и генных панелей в сравнении с КС непрерывно растет. И все больше накапливается доказательств об избыточности валидации во многих случаях. Так, валидация КС «высококачественных вариантов» NGS (глубина прочтения не менее 20x, представленность варианта не менее 20 %, высокая оценка качества вызова варианта) уже не считается необходимой [1]. А итоговое увеличения стоимости исследования, трудозатрат и сроков выдачи заключений при валидации всех результатов NGS с помощью КС приводит скорее к снижению эффективности тандема NGS+КС. При экзомных исследованиях валидацию потенциально каузативных вариантов часто рекомендуется проводить выборочно, например, в случае небольших найденных инсерций/делеций, а также однонуклеотидных вариантов, которые по результатам биоинформатического анализа имеют качество варианта ниже порогового [2]. Однако численные значения таких пороговых значений в литературных данных приводятся редко, поскольку предполагается индивидуальный их расчет в каждой лаборатории (например, [3]: 5x, представленность варианта от 35 %, значения QUAL от 139). В отличие от результатов экзомных и таргетных NGS исследований, информация о сравнении результатов полногеномного секвенирования (WGS) и валидации КС в литературе представлена очень ограниченно. В отличие от подходов таргетного секвенирования, методология проведения WGS с использованием безамплификационного (PCR-free) подхода предполагает максимальную однородность результатов секвенирования и позволяет получить данные, которые априори имеют наибольшую равномерность охвата по всей длине генома, наименьшее число экспериментальных артефактов и меньшие

требования по необходимой глубине прочтений [4]. Постоянное увеличение объема и качества WGS данных ставит вопрос о необходимости обновления рекомендаций по применению КС в качестве валидационного инструмента. Так, в рекомендациях [5] 2022 г. акцент делается на необходимости позитивной переоценки использования WGS при диагностике и обновлении ранее опубликованных руководств по проведению NGS с указанием, что варианты, соответствующие определенным внутри конкретной лаборатории критериям и показателям качества обработки данных, могут не требовать дополнительной валидации. Целью нашего исследования явилось обобщение внутрилабораторных критериев качества, по достижении которых валидация варианта с помощью КС может считаться избыточной или подходить близко к этому порогу.

Методы и алгоритмы: Всю экспериментальную работу и биоинформатический анализ данных WGS и КС проводили на базе собственной лаборатории. Пробоподготовку образцов для WGS на платформах DNBseq-T7/G400 (MGI) в режиме PE150 проводили по безамплификационному PCR-free протоколу с предварительной ферментативной фрагментацией. Валидацию всех вариантов с помощью КС выполняли на секвенаторе ABI3500 по стандартным протоколам производителя.

Результаты: На данный момент лаборатория провела наибольшее в РФ число WGS, однако для анализа данных мы использовали 20K+ образцов, из которых ~4500 составляли образцы с выявленными орфанными/редкими вариантами и наследственными опухолевыми синдромами, валидированными по методу Сэнгера. Из валидированных КС образцов: 97.9 % были подтверждены рутинным образом (результаты КС полностью совпали с результатами NGS), 1.1 % требовали дополнительного уточнения и экспертного анализа, в результате которого при валидации было получено подтверждение (достигнуто согласование данных NGS и результатов КС), 1 % составили технические отказы в сложных областях. Полученные результаты позволили сформулировать следующие критерии для возможности непроведения валидации вариантов с помощью КС: глубина прочтения варианта 10x и выше; представленность аллельного варианта 15–20 % и выше; вариант располагается в области, не имеющей псевдогенов; в ближайшем окружении к валидируемому варианту отсутствуют множественные и сложные для амплификации и аннотации инсерции/делеции, протяженные тандемные и однонуклеотидные повторы.

Выводы: 1. Пороговое значение качества вариантов зависит от индивидуальных деталей алгоритма и биоинформатического пайплайна. 2. Валидация методом Сэнгера оказалась полезной для ~1–2 % наиболее сложных вариантов (например, инделов до 30 п. н., если дополнительные варианты расположены в непосредственной близости от валидированного варианта, областей с повторами, GC-регионов, псевдогенов, вариантов с покрытием <10x), однако большинство из этих вариантов все еще можно вызвать *in silico* с помощью дополнительной биоинформатической обработки данных NGS. 3. Однако КС по-прежнему сохраняет актуальность для валидации самих клинических кейсов, т. е. подтверждения действительной каузативности найденных в результате биоинформационной обработки вариантов и проведения сегрегационного анализа. Это приводит к тому, что наиболее эффективным способом более информативного и обширного сегрегационного анализа видится проведение WGS с пониженной глубиной прочтения (~10x) для родственников в дополнение к стандартному 30x секвенированию пробанда.

Summary of intralaboratory quality criteria for genomic variants validated by whole-genome sequencing results

Antonenko A.[1]*, Belov R.[1], Revkova M.[1], Domoratskaya E.[1], Bartcic A.[1], Panferova A.[1], Gubona M.[1], Ulanova P.[1], Dibirova H.[1], Sokolova N.[1], Doroshuk N.[1], Makarova M.[1], Groznova O.[2], Krinitsina A.[1], Belenikin M.[1]**

[1] *Evogen LLC, Moscow, Russia*

[2] *Russia Veltischev Research and Clinical Institute for Pediatrics and Pediatric Surgery of the Pirogov Russian National Research Medical University, Moscow, Russia*

** antonenko_a.n@mail.ru; ** evogenlab@yandex.ru*

Key words: whole-genome sequencing, WGS; validation of NGS variants; Sanger sequencing

Motivation and Aim: The rapid development of high-throughput sequencing (NGS) technologies, bioinformatics algorithms, and NGS data processing is associated with the need to validate identified potential causative variants using alternative methods. The most common way to validate the identified variants is by capillary sequencing (CS) using the Sanger method, as this method is considered the so-called 'gold standard' for clinical research. The number of scientific articles comparing NGS results for exomes and gene panels versus CS is continuously increasing. And evidence is accumulating about the redundancy of validation in many cases. Thus, CS validation of 'high-quality variants' of NGS (read depth of at least 20x, variant representation of at least 20 %, high quality score of variant calling) is no longer considered necessary [1]. And the resulting increase in the cost of the study, labour costs and time of issuing conclusions when validating all NGS results with the help of CS leads rather to a decrease in the efficiency of the NGS+CS tandem. In exome studies, validation of potentially causative variants is often recommended to be performed selectively, for example, in the case of small insertions/deletions found, as well as single-nucleotide variants that, according to the results of bioinformatic analysis, have a variant quality below the threshold [2]. However, the numerical values of such thresholds are rarely reported in the literature, as they are supposed to be calculated individually in each laboratory (e. g. [3]: 5x, variant representation from 35 %, QUAL values from 139). Unlike the results of exome and targeted NGS studies, there is very limited information in the literature on the comparison of the results of whole genome sequencing (WGS) and CS validation. In contrast to target sequencing approaches, the methodology of WGS using a non-amplification (PCR-free) approach implies maximum homogeneity of sequencing results and allows obtaining data that a priori have the greatest uniformity of coverage over the entire genome length, the lowest number of experimental artefacts and lower requirements for the necessary depth of reads [4]. The continuous increase in the volume and quality of WGS data raises the question of the need to update the recommendations for the use of CS as a validation tool. For example, the 2022 guidelines [5] emphasise the need to positively reassess the use of WGS in diagnostics and update previously published NGS guidelines to state that variants that meet laboratory-specific criteria and quality indicators for data processing may not require additional validation. The aim of this research was to summarise the intra-laboratory quality criteria at which validation by CS of a WGS-derived variant may be considered redundant or close to this threshold.

Methods and Algorithms: All experimental work and bioinformatics analyses of WGS and CS data were performed in our own laboratory. Sample preparation for WGS on DNBseq-T7/G400 (MGI) platforms in PE150 mode was performed using a non-amplification PCR-free protocol with preliminary enzymatic fragmentation. Validation of all variants by CS was performed on an ABI3500 sequencer according to the producer's standard protocols.
Results: To date, the laboratory has performed the largest number of full genome sequencing in the Russian Federation, but we used 20K+ samples for data analysis, of which ~4500 were samples with identified orphan/rare variants and inherited tumour syndromes validated using the Sanger method. Of the CS validated samples: 97.9 % were validated routinely (CS results were in complete agreement with NGS results), 1.1 % required additional clarification and expert analysis resulting in validation (agreement between NGS data and CS results was achieved), 1 % were technical failures in complex areas. The results obtained allowed us to formulate the following criteria for the possibility of not validating variants using CS: variant read depth – 10x and higher; allelic variant representation 15–20 % and higher; the variant is located in a region without pseudogenes; there are no multiple and difficult for amplification and annotation insertions/deletions, extended tandem and single nucleotide repeats in the immediate vicinity of the variant to be validated.
Conclusion: (1) The threshold for variant quality depends on the individual details of the algorithm and bioinformatic pipeline. (2) Sanger validation has proven useful for ~1–2 % of the most difficult variants (e. g. indels up to 30 bp if additional variants are located in close proximity to the validated variant, regions with repeats, GC regions, pseudogenes, variants with <10x coverage), but most of these variants can still be induced in silico with additional advanced bioinformatic NGS data processing. (3) However, CS still remains relevant for validating the clinical cases themselves, i. e. confirming the actual causality of the variants found by bioinformatics processing and performing segregation analysis. This leads us to believe that the most effective way to perform a more informative and extensive segregation analysis is to perform WGS with reduced read depth (~10x) for relatives in addition to the standard 30x sequencing of the proband.

Список литературы/References

1. Arteche-Lopez A. et al. Sanger sequencing is no longer always necessary based on a single-center validation of 1109 NGS variants in 825 clinical exomes. Sci Rep. 2021;11:5697. doi 10.1038/s41598-021-85182-w
2. Strom S.P. et al. Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. Genet Med. 2014;16(7):510-515. doi 10.1038/gim.2013.183
3. Bauer P. et al. Development of an evidence-based algorithm that optimizes sensitivity and specificity in ES-based diagnostics of a clinically heterogeneous patient population. Genet Med. 2019;21(1):53-61. doi 10.1038/s41436-018-0016-6
4. Meynert A.M. et al. Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics. 2014;15:247. doi 10.1186/1471-2105-15-247
5. Souche E. et al. Recommendations for whole genome sequencing in diagnostics for rare diseases. Eur J Hum Genet. 2022;30:1017-1021. doi 10.1038/s41431-022-01113-x

doi 10.18699/bgrs2024-4.2-02

Популяционная транскриптомика преэклампсии

Бабовская А.А.*, Трифонова Е.А., Степанов В.А.

Научно-исследовательский институт медицинской генетики, Томский НИМЦ РАН, Томск, Россия
** anastasia.babovskaya@medgenetics.ru*

Ключевые слова: плацента; преэклампсия; популяции; транскриптом; высокопроизводительное секвенирование

Мотивация и цель: Преэклампсия (ПЭ) является одним из наиболее тяжелых осложнений беременности, она оказывает негативное влияние как на состояние здоровья самой беременной, так и на состояние плода и новорожденного. Многочисленные исследования указывают на многофакторную природу ПЭ, со значимой ролью в ее формировании генетических факторов, однако причины возникновения ПЭ до настоящего времени остаются неизвестными. Считается, что основополагающими в патогенезе ПЭ являются нарушения при формировании плаценты, а вероятный механизм таких нарушений заключается в дисрегуляции экспрессии генов, участвующих в процессах плацентации в норме. Стоит отметить, что неоднократно сообщалось как о разной частоте возникновения ПЭ в популяциях человека, так и о наличии этнических особенностей в течении беременности в целом. В связи с этим исследования, направленные на изучение экспрессии генов при физиологической беременности (ФБ) и ПЭ в различных популяциях, несомненно, являются актуальными. Таким образом, целью настоящего исследования были изучение транскриптомного профиля децидуальных клеток (ДК) человека и оценка его межпопуляционной вариабельности при физиологической беременности и преэклампсии в популяциях России.
Методы и алгоритмы: Осуществлен полнотранскриптомный анализ (RNA-seq) ДК плаценты, полученных с помощью технологии лазерной микродиссекции криоконсервированных биоптатов материнской части плаценты от русских и бурятских женщин с ФБ и ПЭ.
Результаты: Анализ дифференциальной экспрессии генов ДК в обследованных группах женщин позволил выявить статистически значимые различия транскриптомного профиля при сравнении здоровых и больных индивидов из популяционных выборок русских и бурят. Установлено, что гены, ассоциированные с ПЭ у русских, сверхпредставлены в процессах метаболизма и активности молекул – производных жирных кислот, ионного транспорта и поляризации мембраны. Физиологическими эффектами изменения экспрессии генов, участвующих в перечисленных процессах, является нарушение водно-солевого обмена, которое рассматривается как значимый фактор развития отеков и артериальной гипертензии – основных патогенетических механизмов преэклампсии. В свою очередь у женщин с ПЭ зафиксировано увеличение атерогенных триглицеридов низкой плотности и снижение концентрации липопротеинов высокой плотности, что способствует атерогенезу и дисфункции спиральных артерий.
У бурятских женщин изменения экспрессионного профиля ДК в результате ПЭ затрагивают сигнальный путь клеточного апоптоза, пути регуляции биосинтеза и

процессов транскрипции, а также процессы дифференцировки костных, хрящевых и эпителиальных клеток, необходимых для нормального роста и развития организма. Изменение баланса между программами пролиферации и апоптоза в эндотелиальных клетках, а также в популяциях трофобласта могут приводить к плацентарной ишемии, нарушению процессов ангиогенеза и инвазии, что в свою очередь будет способствовать развитию преэклампсии. Кроме того, нарушение экспрессии генов, ответственных за дифференцировку костных, хрящевых и эпителиальных клеток, вероятно, может рассматриваться как один из патогенетических факторов задержки роста плода.

Полученные оценки вариабельности в представленных выборках свидетельствуют о том, что доля межпопуляционных различий среди индивидов с физиологической беременностью составляет 6 %. В то же время аналогичный показатель межгрупповой вариабельности (Nst) в группах с ПЭ составил всего 1 %, такая тенденция может быть в результате того, что молекулярные механизмы болезни приводят к частичной элиминации популяционной специфики транскриптомов. Интересно отметить, что среди индивидов в группах с ПЭ доля внутрипопуляционной изменчивости была меньше (Nit = 46 %), чем данный показатель в группе с физиологической беременностью (Nit = 67 %). Таким образом, в работе показано, что развитие патологического процесса приводит к снижению межиндивидуальной и межпопуляционной вариабельности, что проявляется в большем сходстве транскриптомных профилей ДК пациенток с ПЭ по сравнению с женщинами с физиологической беременностью.

Выводы: Впервые на уровне отдельных клеток плацентарной ткани была проведена оценка внутри- и межпопуляционной вариабельности полногеномной экспрессии генов. Эти данные имеют большое значение для изучения механизмов здоровой плацентации, а обнаруженные общие и специфические функциональные процессы могут лечь в основу дальнейших исследований популяционных особенностей акушерских осложнений.

Финансирование: Исследование выполнено за счет средств государственного задания по теме ФНИ № 122020200083-8.

Population transcriptomics of preeclampsia

Babovskaya A.A.*, Trifonova E.A., Stepanov V.A.

Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia

** anastasia.babovskaya@medgenetics.ru*

Key words: placenta; preeclampsia; populations; transcriptome; high throughput sequencing

Motivation and Aim: Preeclampsia (PE) is one of the most severe complications of pregnancy; it has a negative impact on both the health of the pregnant woman and the fetus. Numerous studies indicate the multifactorial nature of PE with a significant role in its formation of genetic factors, but the causes of PE still remain unknown. It is believed that disturbances during the formation of the placenta are fundamental in the pathogenesis of PE. The probable mechanism of such disturbances is the dysregulation of the expression of genes involved in normal placentation processes. It is worth noting that it has been reported about the different frequency of PE in human populations and in general about the presence of ethnic characteristics during pregnancy. In this regard,

studies aimed at studying gene expression during normal pregnancy (NP) and PE in various populations are undoubtedly relevant. Thus, the purpose of this study was to study the transcriptomic profile of human decidual cells (DCs) and assess its inter-population variability during NP and PE in Russian populations.

Methods and Algorithms: We provide an analysis of full transcriptome (RNA-seq) of placental DCs obtained using laser microdissection of the maternal part of placenta from Russian and Buryat women with NP and PE.

Results: Analysis of the differential expression of DCs genes in the examined groups of women made it possible to identify statistically significant differences in the transcriptomic profile when comparing individuals with NP and PE from population samples of Russians and Buryats. It has been established that genes associated with PE in Russians women are overrepresented in the processes of metabolism and activity of fatty acid molecules, ion transport and membrane polarization. The physiological effects of changes in genes involved in the listed processes is a violation of water-salt metabolism which is considered as a significant factor in the development of edema and arterial hypertension – the main pathogenetic mechanisms of preeclampsia. In turn, in women with PE was recorded an increase in atherogenic low-density triglycerides and a decrease in the concentration of high-density lipoproteins, which contributes to atherogenesis and dysfunction of the spiral arteries.

In Buryat women, changes in the expression profile of DCs as a result of PE affect the signaling pathway of cell apoptosis, pathways for regulating biosynthesis and transcription processes, as well as the processes of differentiation of bone, cartilage and epithelial cells necessary for normal growth and development of the body. Changes in the balance between proliferation and apoptosis programs in endothelial cells as well as in trophoblast can lead to placental ischemia, disruption of angiogenesis and invasion, which in turn will contribute to the development of preeclampsia. In addition, disruption of the expression of genes responsible for the differentiation of bone, cartilage and epithelial cells can probably be considered as one of the pathogenetic factors of fetal growth retardation.

The obtained estimates of variability in the presented samples indicate that the share of inter-population differences among individuals with normal pregnancy is 6 %. At the same time this indicator of inter-population variability (Nst) in groups with PE was only 1 %; this trend may be due to the fact that the molecular mechanisms of the disease lead to partial elimination of the population specificity of transcriptomes. It is interesting to note that among individuals in the groups with PE the proportion of inter-individual variability was less (Nit = 46 %) than this indicator in the group with normal pregnancy (Nit = 67 %). Thus, the work shows that the development of the pathological process leads to a decrease in inter-individual and inter-population variability, which is manifested in greater similarity of transcriptomic profiles of DCs in patients with PE compared to women with normal pregnancy.

Conclusion: For the first time in single decidual cells of placenta was assessed inter-individual and inter-population variability of genome-wide gene expression. These data are of great importance for studying the mechanisms of healthy placentation and the discovered general and specific functional processes can form the basis for further studies of the population characteristics of obstetric complications.

Funding: The research was carried out at the expense of the state assignment (Basic scientific research No. 122020200083-8).

doi 10.18699/bgrs2024-4.2-03

Выявление дифференциально метилированных при бронхиальной астме и туберкулезе легких генов (реанализ данных)

Бабушкина Н.*, Огородников С., Зарубин А., Брагина Е.

Научно-исследовательский институт медицинской генетики, Томский НИМЦ РАН, Томск, Россия
** nad.babushkina@medgenetics.ru*

Ключевые слова: астма; туберкулез; метилирование ДНК

Мотивация и цель: Феномен дистропии, характеризующийся редким сочетанием на фенотипическом уровне двух патологий у одного индивида, сложен для изучения. Ряд исследований свидетельствует о возможной дистропии аллергических болезней и туберкулеза (см., например, [1]). Так, показано, что диагноз и астмы, и туберкулеза у одного индивида встречается в разных популяциях с частотой не более чем 1–5 % [1, 2]. Иммунологические характеристики астмы и туберкулеза показывают, что в основе дистропии «астма-туберкулез» может лежать разнонаправленная поляризация иммунного ответа с преимущественной активностью Th2 при астме и Th1 при туберкулезе [3, 4]. Изучение изменчивости регуляции генов при этих патологиях дает дополнительную информацию, важную для понимания феномена дистропии «астма-туберкулез».

Методы и алгоритмы: Проведен реанализ двух наборов данных метилирования ДНК, представленных в общедоступном репозитории функциональных геномных данных Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). GSE118469 получен при изучении туберкулеза у жителей Тайваня: пациенты с туберкулезом ($n = 12$) без сопутствующих инфекционных патологий (в том числе ВИЧ) и злокачественных новообразований, контрольная группа – здоровые индивиды ($n = 6$) [5]. Из набора данных GSE104471 была использована информация о профилях метилирования у 12 пациентов с аллергической бронхиальной астмой и 12 здоровых индивидов без астмы. Все участники этого исследования являются взрослыми некурящими европеоидами неиспаноязычного происхождения [6]. Анализ метилирования в обоих исследованиях проведен на чипах Infinium HumanMethylation450K BeadChip v1.2 [5, 6]. В реанализ из обоих наборов данных взята информация о метилировании ДНК из мононуклеаров периферической крови.

Реанализ проводили с использованием языка программирования R-4.3.3 (R Foundation, Австрия) в среде разработки RStudio-2021.09.1-372 (Rstudio PBC, США). Из наборов данных GSE118469 и GSE104471 был создан общий массив данных с информацией о метилировании ДНК как у больных с астмой и с туберкулезом, так и в контрольных образцах. Далее с применением программного пакета minfi [7] полученный массив подвергали квантильной нормализации, удаляли слабо сработавшие зонды и пробы полиморфных вариантов с частотой встречаемости минорного аллеля больше 5 %, исключали данные с половых хромосом. Поиск дифференциально метилированных у больных астмой и

туберкулезом CpG-сайтов осуществляли при помощи limma [8]. Для исключения межэтнических различий этнос использовался в качестве дополнительной группирующей переменной, после получения взвешенной оценки по каждому из признаков объясняемая ими дисперсия вычиталась. Сходный подход использовали и при деконволюции по клеточному составу, проведенной с использованием программного пакета EpiDISH [9].

Результаты: В результате был получен список CpG-сайтов, дифференциально метилированных (ДМС) у больных с двумя патологиями – астмой и туберкулезом. Поправку на множественные сравнения (adj_$P < 0.05$) прошли 6653 сайта; из них $|\Delta\beta| > 20$ % выявляется для 62 сайтов. Прошедшие оба фильтра сайты относятся к 53 генам (ДМГ).

Анализ сверхпредставленности (over-representation analysis, ORA)) с помощью Web Gestalt (https://www.webgestalt.org/) показал вовлеченность этих генов в биологические процессы (в категории GO Biological Processes), связанные с иммунным ответом (сигнальный путь рецепторов клеточной поверхности; сигнальный путь, опосредованный антигенным рецептором; позитивная регуляция CD8+; ответ на цитокины; продукция цитокинов; адаптивный иммунный ответ; дифференцировка альфа-бета-Т-клеток, иммунный ответ, активация альфа-бета-Т-клеток), три из них проходят поправку FDR: активация альфа-бета-Т-клеток (GO:0046631, $p = 0.010031$), иммунный ответ (GO:0006955, $p = 0.010298$), дифференцировка альфа-бета-Т-клеток (GO:0046632, $p = 0.010298$). В эти три пути вовлечены в общей сложности 17 генов (*CX3CR1*, *CD247*, *CYFIP1*, *ETS1*, *FASLG*, *IGSF6*, *IL19*, *LY9*, *HLA-E*, *PRDM1*, *PVRIG*, *RFTN1*, *RORA*, *RUNX1*, *SYK*, *SURF4*, *ZC3HAV1*), причем 5 из них (*LY9*, *PRDM1*, *RORA*, *RUNX1*, *SYK*) участвуют во всех трех биологических процессах.

Согласно результатам обогащения, полученным с помощью ресурса Metascape (metascape.org), наиболее представленными процессами являются также активация альфа-бета-Т-клеток (GO:0046631) и регуляция активации альфа-бета-Т-клеток (GO:0046634), цитокиновый сигналинг в иммунной системе (Reactome Gene Sets: R-HSA-1280215), цитоксичность, опосредованная NK-клетками (KEGG Pathway: hsa04650), путь дифференцировки клеток Th17 (WikiPathways: WP5130). Анализ ассоциаций с патологиями и физиологическими состояниями (по Metascape) показал связь выявленных ДМГ с количеством эозинофилов и лимфоцитов, средним уровнем гемоглобина, заболеваниями иммунной системы (в том числе – синдром Бехчета, васкулит, первичный синдром Шегрена, иммунная тромбоцитопеническая пурпура, обыкновенная волчанка, красная волчанка, дискоидная красная волчанка, болезнь Либмана–Сакса, астма у взрослых, реакция «трансплантат против хозяина») и связанными с ними заболеваниями (лимфаденопатия, моноклональные гаммапатии), а также с карциномой яичников, лимфоидным лейкозом, венозной тромбоэмболией, отеком суставов. Помимо этого, выявленные ДМГ ассоциированы в первую очередь с CD56+ NK-клетками, селезенкой и тимусом (по результатам анализа в Metascape).

Ресурс FUMA GWAS (https://fuma.ctglab.nl/) указывает на наличие ассоциаций с различными патологиями/физиологическими показателями для 19 из выявленных ДМГ (*AFF3*, *ANKH*, *APBA2*, *CX3CR1*, *ETS1*, *FASLG*, *GIMAP4*, *HLA-E*, *IKZF1*, *IL19*, *MPHOSPH9*, *RORA*, *SLA*, *SPATA13*, *SYK*, *TG*, *UBAC2*, *ZC3HAV1*, *CD247*). Так, выявлены ассоциации с эозинофилами (количество/процентное содержание среди белых клеток), тромбокрит, соотношение альбумин-глобулин; темперамент;

рассеянный склероз, эрозия костей при ревматоидном артрите, аутоиммунные заболевания щитовидной железы, применение медикаментов (препараты для щитовидной железы, адренергические средства). Кроме того, показаны ассоциации с астмой: как в целом (*CD247*, *FASLG*, *AFF3*, *ANKH*, *IKZF1*, *TG*, *ETS1*, *UBAC2*, *RORA*), так и с аллергической формой (*FASLG*, *IKZF1*, *ETS1*, *RORA*).
Следует отметить, что распределение ДМС не симметрично – все 62 сайта с $|\Delta\beta|>20$ % имеют отрицательное значение. При расчете $\Delta\beta$ проводилось вычитание среднего уровня метилирования у больных с туберкулезом из среднего уровня метилирования у больных с астмой. Поэтому можно утверждать, что у пациентов с астмой происходит относительное снижение уровня метилирования ДМС, что может свидетельствовать об усилении активности ДМГ. Поскольку связь между уровнем экспрессии и уровнем метилирования нелинейна и к тому же различается в зависимости от расположения CpG-сайтов в пределах гена, напрямую говорить об увеличении экспрессии ДМГ будет опрометчивым.
Выводы: Полученные нами данные свидетельствуют о том, что при развитии таких дистропных патологий, как астма и туберкулез, происходит изменение активности прежде всего именно тех генов, которые вовлечены как в развитие иммунного ответа в целом, так и в активацию и дифференциацию альфа-бета-Т-клеток. Полученные данные полностью укладываются в гипотезу поляризации Т-клеточного иммунного ответа при дистропии «астма-туберкулез».
Финансирование: Работа выполнена в рамках государственного задания Министерства науки и высшего образования (№ 122020300041-7).

Identification of differentially methylated genes in bronchial asthma and pulmonary tuberculosis (re-analysis of data)

Babushkina N.*, Ogorodnikov S., Zarubin A., Bragina E.
Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia
** nad.babushkina@medgenetics.ru*

Key words: asthma; tuberculosis; DNA methylation

Motivation and Aim: The phenomenon of dystropyis characterized by a rare combination at the phenotypic level of two pathologies in one individual and is difficult to study. A number of studies indicate a possible dystropy of allergic diseases and tuberculosis (see, for example, [1]). Thus, it has been shown that a combination of both asthma and tuberculosis in one individual occurs in different populations with a frequency no more than 1–5 % [1, 2]. The immunological characteristics of both diseases show that the basis of the dystropy “asthma-tuberculosis” may be multidirectional polarization of the immune response with predominant activity of Th2 in asthma and Th1 in tuberculosis [3, 4]. Studying the variability of gene regulation in these pathologies provides important information for understanding the phenomenon of “asthma-tuberculosis” dystropy.
Methods and Algorithms: Two DNA methylation datasets were re-analyzed from the Gene Expression Omnibus public functional genomic data repository (GEO, https://www.ncbi.nlm.nih.gov/geo/). Microarray dataset GSE118469 was obtained in a study of tuberculosis in residents of Taiwan: patients with tuberculosis (n = 12) without concomitant infectious (including HIV) and malignant neoplasms, the control group – healthy individuals (n = 6) [5]. We analized information on methylation profiles of

12 patients with allergic bronchial asthma and 12 healthy individuals without asthma from GSE104471 dataset. All participants are non-Hispanic white non-smoking adults [6]. Infinium HumanMethylation450K BeadChip v1.2 was used for methylation analysis in both studies [5, 6]. In the re-analysis, DNA methylation data from PBMCs are taken from both datasets.

Re-analysis was carried out using the R-4.3.3 programming language (R Foundation, Austria) in the RStudio-2021.09.1-372 development environment (Rstudio PBC, USA). Data from datasets GSE118469 and GSE104471 were integrated in common dataset with information on DNA methylation in both patients with asthma and tuberculosis, and in control samples. Next, using the minfi software package [7], the obtained dataset was subjected to quantile normalization. Weakly working probes and samples of polymorphic variants with a minor allele frequency of more than 5 % were removed and data from sex chromosomes were excluded. The differentially methylated CpG sites in patients with asthma and tuberculosis were searched using limma [8]. To exclude interethnic differences, ethnicity was used as an additional grouping variable; after obtaining a weighted estimate for each of the attributes, the variance explained by them was subtracted. A similar approach was used for deconvolution by cellular composition, carried out using the EpiDISH software package [9].

Results: As a result, a list of differentially methylated CpG sites (DMS was obtained in patients with two pathologies – asthma and tuberculosis. Adjustment for multiple comparisons (adj_P < 0.05) passed 6653 sites; of these, $|\Delta\beta| > 20$ % are identified for 62 sites. The sites that passed both filters belong to 53 genes (DMG).

Over-Representation analysis (ORA) using Web Gestalt (https://www.webgestalt.org/) showed the involvement of these genes in biological processes (in the GO Biological Processes category) associated with the immune response; activating cell surface receptor signaling pathway; antigen receptor-mediated signaling pathway; positive regulation of CD8-positive; alpha-beta T cell activation; response to cytokine; cytokine production; adaptive immune response; alpha-beta T cell differentiation; immune response), three of them undergo FDR correction: alpha-beta T cell activation (GO:0046631, $p = 0.010031$), immune response (GO:0006955, $p = 0.010298$), alpha-beta T cell differentiation (GO :0046632, $p = 0.010298$). A total of 17 genes are involved in these three pathways (*CX3CR1*, *CD247*, *CYFIP1*, *ETS1*, *FASLG*, *IGSF6*, *IL19*, *LY9*, *HLA-E*, *PRDM1*, *PVRIG*, *RFTN1*, *RORA*, *RUNX1*, *SYK*, *SURF4*, *ZC3HAV1*), and 5 of them (*LY9*, *PRDM1*, *RORA*, *RUNX1*, *SYK*) are involved in all three biological processes. According to the enrichment results obtained using the Metascape resource (metascape.org), the most represented processes are also alpha-beta T cell activation (GO:0046631), as well as regulation of alpha-beta T cell activation (GO:0046634), Cytokine Signaling in Immune system (Reactome Gene Sets: R-HSA-1280215), Natural killer cell mediated cytotoxicity (KEGG Pathway: hsa04650), Th17 cell differentiation pathway (WikiPathways: WP5130). Analysis of associations with pathologies and physiological conditions (according to Metascape) showed the connection of identified DMGs with Eosinophil and Lymphocyte count, Mean Hemoglobin, Immune System Diseases (including Behcet Syndrome, Vasculitis, Primary Sjögren's syndrome, Immune thrombocytopenic purpura, Lupus Vulgaris, Lupus Erythematosus, Discoid, Lupus Erythematosus, Libman-Sacks Disease, Adult onset asthma, Graft-vs-Host Disease) and related diseases (Lymphadenopathy, Monoclonal Gammapathies), as well as Ovarian Carcinoma, Lymphoid leukemia, Venous Thromboembolism, Joint swelling. In

addition, the identified DMGs are associated primarily with Cell-specific: CD56+ NKCells, spleen and thymus (based on Metascape analysis).
The FUMA GWAS resource (https://fuma.ctglab.nl/) indicates associations with various pathologies/physiological indicators for 19 of the revealed DMGs (*AFF3*, *ANKH*, *APBA2*, *CX3CR1*, *ETS1*, *FASLG*, *GIMAP4*, *HLA-E*, *IKZF1*, *IL19*, *MPHOSPH9*, *RORA*, *SLA*, *SPATA13*, *SYK*, *TG*, *UBAC2*, *ZC3HAV1*, *CD247*). Thus, associations were identified with Eosinophil (counts/percentage of white cells), plateletcrit, Albumin-globulin ratio; Temperament; Multiple sclerosis, Bone erosion in rheumatoid arthritis, Autoimmune thyroid disease, Medication use (thyroid preparations, adrenergics). In addition, associations with asthma are shown: both in general (*CD247*, *FASLG*, *AFF3*, *ANKH*, *IKZF1*, *TG*, *ETS1*, *UBAC2*, *RORA*) and with the allergic form (*FASLG*, *IKZF1*, *ETS1*, *RORA*).
Of note that the DMS distribution is not symmetric – all 62 sites with $|\Delta\beta|>20$ % have a minus value. When calculating $\Delta\beta$, the average methylation level in patients with tuberculosis was subtracted from the average methylation level in patients with asthma. Therefore, it can be argued that in patients with asthma there is a relative decrease in the level of DMS methylation, which may indicate increased DMG activity. Since the relationship between the level of expression and the level of methylation is not linear, moreover, it differs depending on the location of CpG sites within the gene, it would be rash to directly talk about an increase in DMG expression.
Conclusion: Our data directly indicate that in development of dystropic pathologies such as asthma and tuberculosis, there is a change in the activity primarily of those genes that are involved both in the development of the immune response in general and in alpha-beta T cell activation and differentiation. The data obtained are fully fit into the hypothesis of polarization of the T-cell immune response in “asthma-tuberculosis” dystropy.
Funding: This work was supported by the Ministry of Science and Higher Education (project No. 122020300041-7).

Список литературы/References

1. Fekih L. et al. Asthme et tuberculose: association bénéfique ou maléfique? [Tuberculosis in patients with asthma]. *Rev Mal Respir*. 2010;27(7):79-684. doi 10.1016/j.rmr.2010.06.010
2. Cernat T. et al. Simoultaneuos occurence of other diseases among prison inmates with tuberculosis. *Curr Health Sci J*. 2010;36(3):143-147
3. Brighenti S., Ordway D.J. Regulation of immunity to tuberculosis. *Microbiol Spectr*. 2016;4(6). doi 10.1128/microbiolspec.TBTB2-0006-2016
4. Hammad H., Lambrecht B.N. The basic immunology of asthma. *Cell*. 2021;184(9):2521-2522. doi 10.1016/j.cell.2021.04.019. Erratum for: *Cell*. 2021;184(6):1469-1485
5. Chen Y.C. et al. Whole genome dna methylation analysis of active pulmonary tuberculosis disease identifies novel epigenotypes: PARP9/miR-505/RASGRP4/GNG12 gene methylation and clinical phenotypes. *Int J Mol Sci*. 2020;21(9):3180. doi 10.3390/ijms21093180
6. Yang I.V. et al. The nasal methylome: a key to understanding allergic asthma. *Am J Respir Crit Care Med*. 2017;195(6):829-831. doi 10.1164/rccm.201608-1558LE
7. Aryee M.J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-1369. doi 10.1093/bioinformatics/btu049
8. Smyth G.K. Limma: linear models for microarray data. In: Gentleman R. (Ed.). Bioinformatics and Computational Biology Solutions Using R and Bioconductor. 2005;397-420. doi 10.1007/0-387-29362-0_23
9. Teschendorff A.E. et al. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinformatics*. 2017;18(1):105. doi 10.1186/s12859-017-1511-5

doi 10.18699/bgrs2024-4.2-04

Феномен сочетания болезней: роль генетических факторов в развитии синтропии и дистропии

Брагина Е.Ю.[1]*, Пузырев В.П.[1, 2]

[1] *Научно-исследовательский институт медицинской генетики, Томский НИМЦ РАН, Томск, Россия*

[2] *Сибирский государственный медицинский университет Минздрава России, Томск, Россия*

* *elena.bragina@medgenetics.ru*

Ключевые слова: феномен сочетания болезней; синтропия; дистропия, гены, SNPs

Мотивация и цель: Более 50 миллионов человек в возрасте 65 лет и старше, составляющих почти половину населения Европы, страдает одновременно двумя и более многофакторными заболеваниями (МФЗ) [1]. В условиях недостаточности знаний о патогенезе и формировании сочетаний МФЗ у человека в ближайшие 20 лет прогнозируется неуклонный рост (до 68 %) численности коморбидных пациентов и, соответственно, затрат ресурсов здравоохранения, направленных на их лечение и профилактику [2, 3]. Одной из причин фенотипических ассоциаций предполагают участие генетических факторов, которые могут привести к развитию множественных фенотипов [4–6] и остаются на данный момент недостаточно изученными. Основной целью является исследование феномо-геномных отношений в развитии некоторых синтропных (наиболее часто сочетающихся) и дистропных (редко сочетающихся) форм МФЗ.

Методы и алгоритмы: В работе изучены молекулярно-генетические особенности для болезней синтропий, включающих распространенные сочетания бронхиальной астмы с другими аллергическими и сердечно-сосудистыми заболеваниями, и дистропии бронхиальной астмы и туберкулеза. Для исследования связей между болезнями использованы: 1) биоинформатические методы для идентификации генов синтропных и дистропных форм болезней, включая реконструкцию и анализ ассоциативных генных сетей и приоритизацию регуляторных полиморфных вариантов (SNPs); 2) молекулярно-генетические методы: генотипирование SNPs с помощью ПЦР-ПДРФ анализа, реал-тайм ПЦР, анализа кривых плавления с высоким разрешением, мультиплексное генотипирование с помощью масс-спектрометрии MALDI TOF, а также полногеномное генотипирование на платформе Illumina 610-Quad; 3) оценка функциональной значимости полиморфных вариантов генов в зависимости от генотипа и вида стимулятора, вызывающего иммунный ответ, с использованием краткосрочных культур мононуклеарных клеток; (4) общепринятые статистические методы.

Результаты: В результате исследования установлено, что сочетание бронхиальной астмы как с другими аллергическими заболеваниями (поллиноз и атопический дерматит), так и с сердечно-сосудистыми заболеваниями (гипертоническая болезнь) представляет собой отдельные фенотипы, отличающиеся по своим молекулярно-генетическим характеристикам от их «изолированных» форм. Выявлено, что ключевые патофизиологические механизмы, лежащие в основе неблагоприятного сочетания аллергических

заболеваний, связаны с генами иммунной системы и генами, регулирующими функционирование эпидермального барьера. Риск развития сопутствующей гипертонической болезни у пациентов с бронхиальной астмой преимущественно связан с полиморфизмом генов *TLR4*, *CAT*, *ANG/RNASE4*, которые определяют важность воспаления, окислительного стресса и неоваскуляризации, а также генов-мишеней лекарств, в том числе *ADRB1*, *ADRB2*, *NR3C1*, *ANXA1*, использующихся для лечения бронхиальной астмы и гипертонической болезни. Установлены SNPs, ассоциированные с риском развития как бронхиальной астмы, так и туберкулеза (*IFNG* (rs2069705), *TNFB* (rs2239704) и *TNFRSF1B* (rs652625)), а также варианты, которые связаны только с одним из заболеваний. В частности, с бронхиальной астмой ассоциированы варианты генов *PIAS3* (rs3760903), *SOCS5* (rs6737848), с туберкулезом – *PIASY* (rs3760903), *CXCL10* (rs56061981 и rs4386624). Для ассоциированных как с астмой, так и с туберкулезом вариантов генов выявлено, что одни и те же генотипы и аллели «общих» для данных патологий генов-кандидатов являются рисковыми/протективными для развития заболеваний. Однако выявлены некоторые функциональные особенности ассоциированных SNPs. В частности, в условиях стимуляции мононуклеарных клеток крови индуктором микробного происхождения установлено выраженное изменение экспрессии гена *TNF* в зависимости от генотипа индивида для полиморфизма rs2239704. Основные молекулярные функции генов, ассоциированных с развитием бронхиальной астмы и туберкулеза, относятся к цитокиновой активности, связыванию с рецепторами и белками, регуляции активности TNF – важной молекулы с точки зрения как риска развития туберкулеза, так и утяжеления симптомов астмы. Это дает основание предположить, что инфекционные и аллергические заболевания представляют собой варианты одного феномена, обусловленного неадекватной работой иммунной системы, основанной на общей генетической составляющей.

Выводы: Таким образом, феномен сочетания болезней представляет самостоятельный интерес для исследований фундаментального характера и становится дополнительным путем выяснения этиологии и патогенеза сложных заболеваний. Исследование медицинских аспектов разного уровня сочетаний клинических фенотипов позволяет оценить пользу/вред взаимоотношений между болезнями, решая задачи трансляционной геномики, в том числе в области диагностики и терапии заболеваний.

Финансирование: Исследование поддержано Министерством науки и высшего образования (№ 122020300041-7).

The diseases connection phenomenon: the role of genetic factors in the development of syntropy and dystropy

Bragina E.Yu.[1]*, Puzyrev V.P.[1, 2]

[1] *Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia*

[2] *Siberian State Medical University, Tomsk, Russia*

* *elena.bragina@medgenetics.ru*

Key words: the diseases connection phenomenon; syntropy; dystropy; genes; SNPs

Motivation and Aim: More than 50 million people aged 65 years and older, accounting for almost half of the European population, suffer simultaneously from two or more multifactorial diseases (MFDs) [1]. Given the lack of knowledge about the pathogenesis and formation of combinations of MFDs, a steady increase (up to 68 %) in the number of comorbid patients and, accordingly, the cost of health care resources aimed at their treatment and prevention is predicted in the next 20 years [2, 3]. One of the reasons for phenotypic associations is assumed to be the involvement of genetic factors, which can lead to the development of multiple phenotypes [4–6] and remain insufficiently studied at the moment. The main goal is to study phenome-genomic relationships in the development of some syntropy (most often combined diseases) and dystropy (rarely combined diseases).

Methods and Algorithms: In this work we studied molecular genetic features for syntropic diseases, including common coexisting of bronchial asthma with other allergic/cardiovascular diseases, and dystropy of bronchial asthma and tuberculosis. We used next methods for study connections between diseases: (1) bioinformatics methods to identify genes of syntropic and dystropic forms of diseases, including reconstruction and analysis of associative gene networks and prioritization of regulatory polymorphic variants (SNPs); (2) molecular-genetic methods: genotyping of SNPs using PCR-RFLP analysis, real-time PCR, high-resolution melting curve analysis, multiplex genotyping using MALDI TOF mass spectrometry, as well as whole-genome genotyping on the Illumina 610-Quad platform; (3) assessment of the functional significance of SNPs depending on the genotype and type of stimulator causing an immune response in short-term cultures of blood mononuclear cells; (4) statistical methods.

Results: As a result of the study, it was established that the combination of bronchial asthma, both with other allergic diseases (hay fever and atopic dermatitis) and cardiovascular diseases (hypertension), represents unlike phenotypes from their "isolated" forms that differ in their genetic characteristics. It has been revealed that the key pathophysiological mechanisms underlying the unfavorable combination of allergic diseases are associated with genes of the immune system and genes regulating the functioning of the epidermal barrier. The risk of developing concomitant arterial hypertension in patients with bronchial asthma is mainly associated with polymorphisms of the *TLR4*, *CAT*, *ANG/RNASE4* genes, which determine the importance of inflammation, oxidative stress and neovascularization, as well as, *ADRB1*, *ADRB2*, *NR3C1*, *ANXA1* genes, that are targets of drugs that are used of bronchial asthma and arterial hypertension treatment. In this study were identified SNPs that influence at the risk of both bronchial asthma and tuberculosis (*IFNG* (rs2069705), *TNFB* (rs2239704) and *TNFRSF1B* (rs652625)), as well as SNPs that are associated only with one of the diseases, in particular, with bronchial asthma associated gene variants *PIAS3* (rs3760903), *SOCS5* (rs6737848); with tuberculosis – *PIASY* (rs3760903), *CXCL10* (rs56061981 and rs4386624). For the associated SNPs with both bronchial asthma and tuberculosis it was revealed that the same genotypes and alleles of candidate genes "common" for these pathologies are risk/protective for the diseases. However, some functional features of the associated SNPs have been identified. In particular, under conditions of stimulation of mononuclear blood cells with an stimulant of microbial origin, a pronounced change in the expression of the *TNF* gene was established depending on the genotype of the individual for the rs2239704. The main molecular functions of genes associated with the development of bronchial asthma and tuberculosis relate to cytokine activity, binding to receptors and proteins, regulation of the activity of

TNF – an important molecule both in terms of the risk of developing tuberculosis and worsening asthma symptoms. This suggests that infectious and allergic diseases are variants of a single phenomenon caused by inadequate functioning of the immune system based on a common genetic component.

Conclusion: Thus, the diseases connection phenomenon is of independent interest for fundamental research and becomes an additional way to elucidate the etiology and pathogenesis of complex diseases. The study of medical aspects of different levels of combinations of clinical phenotypes makes it possible to assess the benefits/harms of relationships between diseases, solving the problems of translational genomics, including in the field of diagnosis and therapy of diseases.

Funding: The study is supported by the State Task of the Ministry of Science and Higher Education (No. 122020300041-7).

Список литературы/References

1. Rijken M. et al. Managing multimorbidity: Profiles of integrated care approaches targeting people with multiple chronic conditions in Europe. *Health Policy*. 2018;122(1):44-52. doi 10.1016/j.healthpol.2017.10.002
2. Babu M., Snyder M. Multi-Omics Profiling for Health. *Mol Cell Proteomics.* 2023;22(6):100561. doi 10.1016/j.mcpro.2023.100561
3. Kingston A. et al. Projections of multi-morbidity in the older population in England to 2035: estimates from the Population Ageing and Care Simulation (PACSim) model. *Age Ageing.* 2018;47(3):374-380. doi 10.1093/ageing/afx201
4. Пузырев В.П. Генетические основы коморбидности у человека. *Генетика*. 2015;51(4):491-502
5. Sánchez-Valle J. et al. Interpreting molecular similarity between patients as a determinant of disease comorbidity relationships. *Nat Commun.* 2020;11(1):2854. doi 10.1038/s41467-020-16540-x
6. Dong G. et al. A global overview of genetically interpretable multimorbidities among common diseases in the UK Biobank. *Genome Med.* 2021;13(1):110. doi 10.1186/s13073-021-00927-6

doi 10.18699/bgrs2024-4.2-05

Ассоциация полиморфного варианта гена *IL6* с тяжестью депрессивных симптомов

Вялова Н.*, Михалицкая Е., Падерина Д.З., Пожидаев И., Рощина О.В.
НИИ психического здоровья, Томский НИМЦ РАН, Томск, Россия
** Natarakitina@yandex.ru*

Ключевые слова: депрессия; интерлейкин 6; полиморфизм гена; тяжесть заболевания

Мотивация и цель: Аффективные расстройства представляют собой гетерогенную группу заболеваний. Их этиология, патофизиологические механизмы и реакция на антидепрессивную терапию до сих пор плохо изучены. Цитокиновая гипотеза депрессии предполагает, что провоспалительные цитокины являются ключевыми факторами, опосредующими нейроэндокринные и нейрохимические изменения при этом заболевании [1]. Повышение уровня провоспалительных цитокинов из-за длительного стресса приводит к формированию хронического нейровоспаления, которое способствует развитию депрессии. При депрессии происходит увеличение уровня экспрессии провоспалительных генов [2]. Гены, кодирующие цитокины, высокополиморфны. Полиморфизмы в промоторной области генов цитокинов могут привести к межиндивидуальной изменчивости транскрипции и экспрессии генов [3, 4], тем самым влияя на патогенез психических расстройств, прогноз их течения и реакцию на фармакотерапию. Целью нашего исследования стал поиск ассоциаций полиморфного варианта rs2069840 гена *IL6* с депрессивными расстройствами и тяжестью депрессивных симптомов, оцененных в ходе терапии.
Методы и алгоритмы: Обследовано 235 пациентов (187 женщин и 48 мужчин) с аффективными расстройствами (МКБ-10: F31, F32, F33), проходивших курс лечения в клинике НИИ психического здоровья Томского НИМЦ. Средний возраст пациентов составил 44 [29; 54] года. Исследование проводилось согласно этическим принципам ведения исследований человека согласно протоколу, утвержденному локальным этическим комитетом НИИ психического здоровья Томского НИМЦ. Оценку тяжести депрессивных симптомов, а также выраженности типичных и атипичных депрессивных симптомов проводили с использованием шкалы SIGH-SAD**,** включающей **17 пунктов шкалы депрессии Гамильтона и 7 пунктов, оценивающих атипичные депрессивные симптомы** до начала и на 14-й и 28-й дни терапии. В качестве контрольной группы обследовали 190 психически и соматически здоровых лиц (96 женщин и 94 мужчины) в возрасте 39 [29.5; 51.0] лет.
Генотипирование полиморфного варианта rs2069840 гена *IL6* проводили методом ПЦР в реальном времени на амплификаторе Applied Biosystems™ QuantStudio™ 5 Real-Time PCR System (Applied Biosystems, США) с использованием наборов «БиоМастер» UDG HS-qPCR Lo-ROX (2x) («Биолабмикс», Россия). Исследование проводили на базе ЦКП «Медицинская геномика» Томского НИМЦ. Ассоциативный анализ проводился с использованием критерия хи-квадрат и уровня значимости менее 0.05. Предварительно проверяли распределение частот

генотипов на соответствие закону Харди–Вайнберга. Для сравнительного анализа клинических данных использовался критерий Манна–Уитни.
Результаты: Различий в частотах встречаемости аллелей и генотипов rs2069840 гена *IL6* между группами больных с аффективными расстройствами и здоровых лиц не выявлено ($p > 0.05$). В группе пациентов с аффективными расстройствами полиморфный вариант rs2069840 гена *IL6* ассоциирован с большей тяжестью депрессии, оцененной на 28-й день терапии по шкале SIHG-SAD для типичных депрессивных симптомов, а также для суммы типичных и атипичных депрессивных симптомов. Носители генотипа GG на 28-й день антидепрессивной терапии имели достоверно более высокий средний суммарный балл для типичных депрессивных симптомов, чем носители генотипов CC и CG (6 [3.25; 10], 4 [2; 6] и 4 [1.75; 7] соответственно, $\chi2 = 6.788$, $p = 0.034$) и более высокий общий средний суммарный балл (10.5 [7; 12], 6 [3; 8] и 6 [3; 10] соответственно, $\chi2=10.856$, $p = 0.004$).
Выводы: В ходе исследования выявлено, что носители генотипа GG полиморфного варианта rs2069840 гена *IL6* на 28-й день антидепрессивной терапии имели большую тяжесть типичных и атипичных депрессивных симптомов.
Финансирование: Исследование поддержано грантом РНФ «Сравнительное изучение роли иммуновоспаления и нейропротекции в патогенезе и клинике аффективных расстройств и алкогольной зависимости» (№ 23-15-00338).

Association of a polymorphic variant of *IL6* gene with the severity of depressive symptoms

Vyalova N.*, Mikhalitskaya E., Paderina D., Pozhidaev I., Roschina O.
Mental Health Research Institute, Tomsk National Research Medical Center, RAS, Tomsk, Russia
** Natarakitina@yandex.ru*

Key words: depression; interleukin 6; gene polymorphism; severity of the disease

Motivation and Aim: Mood disorders are a heterogeneous group of diseases. Their etiology, pathophysiological mechanisms and response to antidepressant therapy are still poorly understood. The cytokine hypothesis of depression proposes that proinflammatory cytokines are key factors mediating neuroendocrine and neurochemical changes in this disease [1]. Increased levels of pro-inflammatory cytokines due to prolonged stress lead to the formation of chronic neuroinflammation, which contributes to the development of depression. In depression, the expression level of pro-inflammatory genes increases [2]. Genes encoding cytokines are highly polymorphic. Polymorphisms in the promoter region of cytokine genes can lead to interindividual variability in gene transcription and expression [3, 4], thereby influencing the pathogenesis of mental disorders, their prognosis, and response to pharmacotherapy. The purpose of our study was to search for associations of the rs2069840 polymorphic variant of *IL6* gene with depressive disorders and the severity of depressive symptoms assessed during therapy.
Methods and Algorithms: We examined 235 patients (187 women and 48 men) with affective disorders (ICD-10: F31, F32, F33) who were treated at the clinic of Mental Health Research Institute of Tomsk National Research Medical Center. The average age

of patients was 44 [29; 54] years. The study was conducted in accordance with the ethical principles of human research according to the protocol approved by the local ethics committee of Mental Health Research Institute of Tomsk National Research Medical Center. The severity of depressive symptoms, as well as the severity of typical and atypical depressive symptoms, was assessed using the SIGH-SAD scale, which includes 17 items of the Hamilton Depression Inventory and 7 items assessing atypical depressive symptoms before the start and on days 14 and 28 of treatment. As a control group, 190 mentally and somatically healthy individuals (96 women and 94 men) aged 39 [29.5; 51.0] years.

Genotyping of the polymorphic variant rs2069840 of *IL6* gene was carried out by real-time PCR on an Applied Biosystems™ QuantStudio™ 5 Real-Time PCR System (Applied Biosystems, USA) using BioMaster UDG HS-qPCR Lo-ROX kits (2x) (Biolabmix, Russia). The study was carried out on the basis of Medical Genomics Shared Use Center of Tomsk National Research Medical Center. Association analysis was carried out using the chi-square test and a significance level of less than 0.05. The distribution of genotype frequencies was preliminarily checked for compliance with the Hardy–Weinberg law. To conduct a comparative analysis of clinical data, the Mann–Whitney test was used.

Results: There were no differences in the frequencies of alleles and genotypes of rs2069840 of *IL6* gene between the groups of patients with affective disorders and healthy individuals ($p > 0.05$). In group of patients with affective disorders, the rs2069840 polymorphic variant of *IL6* gene is associated with greater severity of depression, assessed by the SIHG-SAD scale on the 28th day of therapy, for typical depressive symptoms, as well as for the sum of typical and atypical depressive symptoms. Carriers of GG genotype on the 28th day of antidepressant therapy had a significantly higher average total score for typical depressive symptoms than carriers of CC and CG genotypes (6 [3.25; 10], 4 [2; 6] and 4 [1.75 ; 7], respectively, $\chi2 = 6.788$, $p = 0.034$) and a higher overall average total score (10.5 [7; 12], 6 [3; 8] and 6 [3; 10], respectively, $\chi2 = 10.856$, $p = 0.004$).

Conclusion: The study revealed that carriers of GG genotype rs2069840 polymorphic variant of *IL6* gene on the 28th day of antidepressant therapy had a greater severity of typical and atypical depressive symptoms.

Funding: The study is supported by the Russian Science Foundation “Comparative study of the role of immunoinflammation and neuroprotection in the pathogenesis and clinic of affective disorders and alcohol addiction” (No. 23-15-00338).

Список литературы/References

1. Köhler C.A. et al. Peripheral cytokine and chemokine alterations in depression: A meta-analysis of 82 studies. *Acta Psychiatr Scand*. 2017;135:373-387
2. Mikhalitskaya E.V. et al. Association of single nucleotide polymorphisms of cytokine genes with depression, schizophrenia and bipolar disorder. *Genes*. 2023;14:1460
3. Martin C. et al. The inflammatory cytokines: molecular biomarkers for major depressive disorder? *Biomark Med*. **2015**;*9*:169-180
4. Gao S.-P. et al. Interleukin-6 genotypes and serum levels in chinese hui population. *Int J Clin Exp Med.* **2014**;*7*:2851-2857

doi 10.18699/bgrs2024-4.2-06

Метилирование генов некодирующих РНК при аневризме и атеросклерозе восходящей аорты

Гончарова И.А.[1]*, Зарубин А.А.[1], Шипулина С.А.[1], Панфилов Д.С.[2], Козлов Б.Н.[2], Назаренко М.С.[1]

[1] *Научно-исследовательский институт медицинской генетики, Томский НИМЦ РАН, Томск, Россия*

[2] *Научно-исследовательский институт кардиологии, Томский НИМЦ РАН, Томск, Россия*

* *irina.goncharova@medgenetics.ru*

Ключевые слова: аневризма и атеросклероз восходящей аорты; некодирующие РНК; метилирование ДНК

Мотивация и цель: Некодирующие РНК (нкРНК) регулируют множество физиологических и патологических процессов, приводящих к развитию различных, в том числе сердечно-сосудистых, заболеваний [1]. Исследования, посвященные функционированию нкРНК при аневризме восходящей аорты, единичны [2]. В немногочисленных работах представлены результаты по метилированию ДНК при атеросклерозе, расслоении или аневризме восходящей аорты [3–6], а также экспрессии генов при сочетании аневризмы аорты и атеросклероза [7]. Однако в данных исследованиях гены нкРНК отдельно не рассматриваются. В связи с этим целью настоящего исследования является изучение особенностей метилирования нкРНК при аневризме восходящей аорты как «изолированной» формы, так и при сочетании с атеросклерозом.

Методы и алгоритмы: Оценка уровня метилирования ДНК выполнена у шести пациентов (мужчины в возрасте 48–64 лет) с аневризмой восходящей аорты (дилатированная часть, нерасширенная проксимальная часть дуги (макроскопически интактная ткань) и атеросклеротическая бляшка, локализованная в проксимальной части дуги аорты). Метилирование ДНК оценивалось с помощью RRBS на приборе Illumina HiSeq1500. Результаты секвенирования обработаны с помощью DRAGEN Bio-IT v.3.9.5 (Illumina) и сопоставлены с геномом человека (сборка GRCh38). Дифференциально метилированными (ДМС) считались CpG-сайты с разницей среднего уровня метилирования между группами образцов FDR < 0.05 и $|\Delta\beta| \geq 0.2$.

Результаты: Из 2 322 700 проанализированных CpG-сайтов между дилатированной при аневризме и интактной тканями аорты различия в уровне метилирования выявлены только для сайтов (chr5:93573113 и chr5:93573147), расположенных в первом интроне гена *NR2F1-AS1*. В дилатированной части аорты наблюдается снижение уровня метилирования данных сайтов относительно интактной ткани ($\Delta\beta = -0.30$ и $\Delta\beta = -0.43$ соответственно).

Между образцами атеросклеротических бляшек и интактной ткани аорты выявлено 480 ДМС, из которых 28 относятся к генам нкРНК. Среди них 8 (28.6 %) локализованы внутри последовательности регуляторных элементов (промотор, энхансер, сайт связывания транскрипционного фактора CTCF) и 15 (51.7 %) гипометилированы в атеросклеротической бляшке аорты.

Для пяти генов нкРНК (*NR2F1-AS1*, *LINC01926*, *LINC01280*, *LINC01770*, *LOC100506691* (*LINC02985*), дифференциально метилированных в атеросклеротических бляшках аорты, ранее показано изменение метилирования в тканях грудной аорты при атеросклерозе без аневризмы [4].

Для ДМС *LINC01926* (chr18:58821123), *LINC01770* (chr1:1433596) и *LINC01280* (chr2:216603507) при аневризме и атеросклерозе аорты показано однонаправленное изменение метилирования с дифференциально метилированными регионами этих генов (chr18:58785792–58786168, chr1:1431925–1432353 и chr2:216566016–216566274) при атеросклерозе аорты без аневризмы [4]: в атеросклеротической бляшке аорты независимо от наличия аневризмы гиперметилированы *LINC01926* и *LINC01280*; гипометилирован *LINC01770*.

Гены *LOC100506691* (*LINC02985*) и *NR2F1-AS1* при атеросклерозе аорты без аневризмы (данные Lacey M. с соавторами [4]) и при сочетании атеросклероза и аневризмы аорты (данные настоящего исследования) показывают разнонаправленное изменение метилирования. Регион chr12:122065085–122065517 гена *LOC100506691* (*LINC02985*) гипометилирован в бляшке аорты без аневризмы ($\Delta\beta = -0.26$) [4], тогда как ДМС chr12:122063660, локализованный в промоторе этого гена, гиперметилирован ($\Delta\beta = 0.55$) при сочетанной патологии. Разнонаправленность метилирования ДНК в данном случае, возможно, объясняется тем, что область промотора и тела гена метилируются по-разному, независимо от наличия изолированной или сочетанной патологии аорты.

В настоящем исследовании при аневризме и атеросклерозе аорты в последовательности гена *NR2F1-AS1* выявлен протяженный гипометилированный регион в 39 024 пары оснований (chr5:93544844–93583868), куда входит 81 CpG-сайт. Среди них два (chr5:93573147 и chr5:93563824) являются ДМС ($\Delta\beta = -0.28$ и $\Delta\beta = -0.37$ соответственно). Вместе с тем при атеросклерозе аорты, не сопровождающемся аневризмой, в гене *NR2F1-AS1* выявлено пять гиперметилированных регионов, в том числе chr5:93571579–93571913 и chr5:93572011–93573146 ($\Delta\beta = 0.25$ и $\Delta\beta = 0.32$ соответственно [4]), в последовательности которых в настоящем исследовании обнаружены гипометилированные сайты.

Кроме этого, выявлены ДМС псевдогенов *YBX3P1* и *SMG1P2*, относящиеся к нкРНК. В последовательности *YBX3P1* находится гиперметилированный в бляшке регион chr16:31569429–chr16:31569442, содержащий четыре CpG-сайта, один из которых является ДМС (chr16:31569442 ($\Delta\beta = 0.29$)). В гене *SMG1P2* выявлен единственный ДМС (chr16:29608442), гипометилированный в бляшке аорты ($\Delta\beta = -0.36$).

Как показывает реанализ результатов экспрессии генов в гладкомышечных клетках восходящей аорты (GSE140947), псевдогены *YBX3P1* и *SMG1P2* являются дифференциально экспрессирующимися при аневризме аорты и атеросклерозе [7]. Наблюдается снижение экспрессии обоих генов в пораженной аорте относительно интактных тканей (log2FoldChange = –0.67 и log2FC = –0.56 соответсвенно) [7].

CpG-сайт chr16:31569442 гена *YBX3P1* локализован в промоторе, и, вероятно, повышение его метилирования приводит к снижению функциональной активность гена. Для гена же *SMG1P2* изменения метилирования сайта chr16:29608442 и экспрессии гена не согласуются: снижение метилирования данного сайта сопровождается и снижением экспрессии гена. CpG-сайт

chr16:29608442 локализован в интроне *SMG1P2* и, вероятно, не влияет на экспрессию данного гена.
Выводы: Некодирующие РНК играют роль в развитии аневризмы восходящей аорты независимо от наличия атеросклеротического поражения сосуда. Маркером развития аневризмы аорты может выступать ген *NR2F1-AS1*, характер метилирования которого различается при сочетанной и изолированных формах атеросклероза и аневризмы восходящей аорты: при изолированной аневризме и в сочетании с атеросклеротическим поражением сосуда ген гипометилирован, а при атеросклерозе без аневризмы аорты – гиперметилирован.
Финансирование: Исследование поддержано грантом РНФ (№ 22-25-00701).

DNA methylation of non-coding RNA genes in aneurysm and atherosclerosis of the ascending aorta

Goncharova I.A.[1]*, Zarubin A.A.[1], Shipulina S.A.[1], Panfilov D.S.[2], Kozlov B.N.[2], Nazarenko M.S.[1]

[1] *Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia*
[2] *Cardiology Research Institute, Tomsk National Research Medical Center, RAS, Tomsk, Russia*
* *irina.goncharova@medgenetics.ru*

Key words: aneurysm and atherosclerosis of ascending aorta; non-coding RNA; DNA methylation

Motivation and Aim: Non-coding RNAs (ncRNAs) control many physiological and pathological processes leading to the development of various medical conditions, including cardiovascular diseases [1]. Research on ncRNAs in ascending aortic aneurysm are scarce [2]. A few studies describe DNA methylation in atherosclerosis, dissection or aneurysm of the ascending aorta [3–6], as well as gene expression in aneurysm in a presence of atherosclerosis [7]. However, these studies do not focus on ncRNA genes. In this regard, the purpose of this study is to characterize DNA methylation of ncRNAs in the ascending aortic aneurysmindividually and in combination with atherosclerosis.
Methods and Algorithms: DNA methylation was measured in six patients (men aged 48–64 years) with an aneurysm of the ascending aorta (dilated part, non-dilated proximal part of the arch (intact tissue) and an atherosclerotic plaque localized in the proximal part of the aortic arch. DNA methylation was determined using RRBS on an Illumina HiSeq1500. Sequencing results were processed using DRAGEN Bio-IT v.3.9.5 (Illumina) and compared to the human genome (GRCh38 assembly). CpG sites with FDR<0.05 and $|\Delta\beta| \geq 0.2$ were considered differentially methylated (DMS).
Results: Of the 2,322,700 CpG sites analyzed between dilated and intact aorta, only two sites (chr5:93573113 and chr5:93573147) were differentially methylated. Both CpG sites are located in the first intron of the *NR2F1-AS1* gene. We observe the decrease in the level of DNA methylation of these CpG sites in the dilated aorta compared to intact tissue ($\Delta\beta = -0.30$ and $\Delta\beta = -0.43$, respectively).
In atherosclerotic plaques and intact aortic tissue, we identified 480 DMS, and 28 of them were related to ncRNA genes. Among them, 8 (28.6 %) are localized within regulatory elements (promoter, enhancer, CTCF transcription factor binding site) and 15 (51.7 %) are hypomethylated in the atherosclerotic aortic plaque.

For five ncRNA genes (*NR2F1-AS1*, *LINC01926*, *LINC01280*, *LINC01770*, *LOC100506691* (*LINC02985*)) differentially methylated in our study, a change in DNA methylation in atherosclerotic thoracic aortic tissue without aneurysm was previously shown [4].

CpG sites of *LINC01926* (chr18:58821123), *LINC01770* (chr1:1433596) and *LINC01280* (chr2:216603507) in aortic aneurysm with atherosclerosis have a unidirectional change of DNA methylation in aortic atherosclerosis without aneurysm (chr18:58785792-587861 68, chr1:1431925-1432353 and chr2:216566016-216566274) [4]: in the aortic atherosclerotic plaque, regardless of the presence of an aneurysm, *LINC01926* and *LINC01280* are hypermethylated and *LINC01770* is hypomethylated.

The *LOC100506691* (*LINC02985*) and *NR2F1-AS1* genes in aortic atherosclerosis without aneurysm (data from Lacey M. et al. [4]) and in a combination of atherosclerosis and aortic aneurysm (present study), show multidirectional changes in DNA methylation. The chr12:122065085-122065517 region of the *LOC100506691* gene (*LINC02985*) is hypomethylated (Δβ= –0.26) in the aortic atherosclerotic plaque without aneurysm [4], while in case of comorbidity the DMS chr12:122063660 in the promoter of this gene is hypermethylated (Δβ=0.55). The multidirectionality of DNA methylation, in this case, may be explained by the fact that the promoter region and the gene body are methylated differently, regardless of the presence of isolated or combined aortic pathology.

In the present study, we identified an extended hypomethylated region of 39,024 base pairs (chr5:93544844-93583868) in aortic aneurysm compared to aortic atherosclerosis. This region includes 81 CpG sites and is located in the *NR2F1-AS1* gene. Among them, two CpG sites (chr5:93573147 and chr5:93563824) are DMS (Δβ = –0.28 and Δβ = –0.37, respectively). At the same time, in atherosclerotic aorta without aneurysm, the *NR2F1-AS1* gene contained five hypermethylated regions including chr5:93571579-93571913 and chr5:93572011-93573146 (Δβ=0.25 and Δβ=0.32, respectively [4]). On the contrary, we found hypomethylated CpG sites in these regions.

Along with that, we identified DMSs related to ncRNAs in the sequence of two pseudogenes *YBX3P1* and *SMG1P2*. The region chr16:31569429-chr16:31569442 of *YBX3P1* is hypermethylated in the aortic atherosclerotic plaque. This region contains four CpG sites, one of them is DMS (chr16:31569442 (Δβ=0.29)). A single DMS (chr16:29608442) located in *SMG1P2* gene is hypomethylated in the aortic atherosclerotic plaque (Δβ = –0.36).

Reanalysis of gene expression in ascending aorta smooth muscle cells (GSE140947) shows that *YBX3P1* and *SMG1P2* pseudogenes are differentially expressed in aortic aneurysm and atherosclerosis [7]. The expression of both genes is decreased in aneurysm compared to intact tissue (log2FoldChange= –0.67 and log2FC= –0.56, respectively) [7]. The CpG site chr16:31569442 of the *YBX3P1* gene is localized in the promoter. It is possible that an increase in its DNA methylation leads to a decrease in the functional activity of the gene. For the *SMG1P2* gene, changes in DNA methylation of the chr16:29608442 site and gene expression are not consistent: a decrease in DNA methylation of this site is accompanied by a decrease in gene expression. The CpG site chr16:29608442 is localized in the intron of *SMG1P2* and probably does not affect the expression of this gene.

Conclusion: Non-coding RNAs play a role in the development of ascending aortic aneurysm regardless of the presence of aortic atherosclerotic lesions. *NR2F1-AS1* gene could be a marker for the development of ascending aortic aneurysmas its DNA

methylation pattern differs in combined and isolated forms of atherosclerosis and ascending aortic aneurysm. In both isolated aortic aneurysm and in combination with atherosclerotic lesion the *NR2F1-AS1* gene is hypomethylated, while in aortic atherosclerosis without aneurysm it is hypermethylated.

Funding: The study is supported by the Russian Science Foundation (No. 22-25-00701).

Список литературы/References

1. Jiang W., Agrawal D.K., Boosani C.S. Non-coding RNAs as epigenetic gene regulators in cardiovascular diseases. *Adv Exp Med Biol*. 2020;1229:133-148. doi 10.1007/978-981-15-1671-9_7
2. Li Y., Yang N. Microarray expression profile analysis of long non-coding RNAs in thoracic aortic aneurysm. *Kaohsiung J Med Sci*. 2018;34(1):34-42. doi 10.1016/j.kjms.2017.09.005
3. Zaina S., Heyn H., Carmona F.J., Varol N., Sayols S., Condom E., Ramírez-Ruz J., Gomez A., Gonçalves I., Moran S., Esteller M. DNA methylation map of human atherosclerosis. *Circ Cardiovasc Genet*. 2014;7(5):692-700. doi 10.1161/CIRCGENETICS.113.000441
4. Lacey M., Baribault C., Ehrlich K.C., Ehrlich M. Atherosclerosis-associated differentially methylated regions can reflect the disease phenotype and are often at enhancers. *Atherosclerosis*. 2019;280:183-191. doi 10.1016/j.atherosclerosis.2018.11.031
5. Pan S., Lai H., Shen Y., Breeze C., Beck S., Hong T., Wang C., Teschendorff A.E. DNA methylome analysis reveals distinct epigenetic patterns of ascending aortic dissection and bicuspid aortic valve. *Cardiovasc Res*. 2017;113(6):692-704. doi 10.1093/cvr/cvx050
6. Chen Y., Xu X., Chen Z., Huang B., Wang X., Fan X. DNA methylation alternation in Stanford- A acute aortic dissection. *BMC Cardiovasc Disord*. 2022;22(1):455. doi 10.1186/s12872-022-02882-5
7. Chen P.Y., Qin L., Li G., Malagon-Lopez J., Wang Z., Bergaya S., Gujja S., Caulk A.W., Murtada S.I., Zhang X., Zhuang Z.W., Rao D.A., Wang G., Tobiasova Z., Jiang B., Montgomery R.R., Sun L., Sun H., Fisher E.A., Gulcher J.R., Fernandez-Hernando C., Humphrey J.D., Tellides G., Chittenden T.W., Simons M. Smooth muscle cell reprogramming in aortic aneurysms. *Cell Stem Cell*. 2020;26(4):542-557.e11. doi 10.1016/j.stem.2020.02.013

doi 10.18699/bgrs2024-4.2-07

Интеграция филостратиграфического и филотранскриптомного анализа в исследовании дифференциально экспрессирующихся генов в раковых тканях

Иванов Р.А.[1]*, Лашин С.А.[1, 2], Афонников Д.А.[1, 2], Матушкин Ю.Г.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

* *ivanovromanart@bionet.nsc.ru*

Ключевые слова: филостратиграфия; филотранскриптомика; дифференциальная экспрессия генов; карциномы

Мотивация и цель: Карциномы представляют собой один из наиболее распространенных видов рака, поражающих различные органы человеческого тела. Понимание эволюционного происхождения и его связи с функциональной ролью генов, связанных с этими заболеваниями, важно не только для осмысления природы рака, но и для разработки новых подходов к его диагностике и лечению. В нашем исследовании мы использовали филостратиграфический и филотранскриптомный анализ для изучения и сравнения эволюционных возрастов генов на различных стадиях развития карцином у человека. Филостратиграфический анализ – метод, направленный на определение эволюционного происхождения генов путем анализа присутствия их ортологов – генов, кодирующих гомологичные белки, разошедшиеся в результате видообразования в геномах разных видов. Такой подход не только выявляет ключевые моменты в эволюции генома, когда наблюдался резкий рост числа новых генов, но и помогает идентифицировать гены, уникальные для специфических организмов. Филотранскриптомика же является методом, объединяющим информацию об эволюционном возрасте генов, полученной с помощью филостратиграфии, и данные об уровне экспрессии генов. Этот подход позволяет исследовать взаимосвязь между давностью возникновения генов и изменениями в их активности в контексте различных физиологических состояний, адаптивных ответов или этапов развития организмов или органов. Целью исследования является выявление генов, дифференциально экспрессирующихся в раковых тканях, эволюционный анализ их возраста и изучение возраста с активностью этих генов в развитии опухолей.

Методы и алгоритмы: Для достижения поставленной цели был использован комплексный подход, сочетающий анализ дифференциальной экспрессии генов, филостратиграфический анализ и филотранскриптомный анализ. Для оценки уровня экспрессии генов и выявления дифференциально экспрессирующихся генов использовался алгоритм DESeq2. Филостратиграфический анализ проводился с использованием веб-сервиса Orthoweb, который позволяет определить эволюционное происхождение генов в единицах индекса филотранскриптомного возраста (Phylostratigraphy Age Index – PAI) на основе анализа присутствия их ортологов в других видов. В качестве порога

идентичности был выбран уровень идентичности белковых последовательностей 0.6, рассчитанный алгоритмом Смита–Ватермана. При помощи Orthoweb также были выполнены оценки индекса изменчивости dN/dS (Divergency Index – DI) при сравнении последовательностей генов с их ортологами в семействе Гоминид. Для оценки уровня экспрессии генов и выявления дифференциально экспрессирующихся генов использовался алгоритм DESeq2. Филотранскриптомный анализ проводился с помощью инструмента myTAI, что позволило оценить динамику экспрессии генов в контексте их эволюционного возраста на различных патологических стадиях карцином в единицах индексов возраста транскриптома (Transcriptome Age Index – TAI) и индекса изменчивости транскриптома (Transcriptome Divergency Index – TDI).

Результаты: В этой работе мы использовали данные по экспрессии генов в тканях печеночной аденокарциномы, светлоклеточной аденокарциномы почек, аденокарциномы кишечника, карциномы молочных желез, аденокарциномы простаты, карциномы щитовидной железы, карциномы уротелия мочевого пузыря, эндометриоидной карциномы тела матки и из соответствующих им образцов здоровых тканей из базы данных TCGA. В каждом из исследуемых типов раковых тканей были выделены опухоли четырех патологических стадий по мере прогрессии опухоли. Филостратиграфический анализ выявил два пика представленности генов: на ранних этапах возникновения многоклеточности и на этапе возникновения позвоночных (PAI = 3 и 6). Филотранскриптомный анализ показал изменения в возрасте транскриптома на разных стадиях развития некоторых опухолей, где в ранних и поздней патологических стадиях развития карцином идет повышение возраста транскриптома в раковых тканях, а в промежуточных – понижение, с особыми отличиями в аденокарциномах кишечника, где наблюдается противоположная картина. Также было обнаружено, что здоровые ткани имеют значимо более высокий возраст транскриптома по сравнению с раковыми, что может свидетельствовать о повышенной активации древних генов при онкологических заболеваниях.

Выводы: Анализ дифференциальной экспрессии генов и их эволюционного возраста в карциномах выявил значительные различия между раковыми и здоровыми тканями, а также специфические изменения в динамике возраста транскриптома на разных стадиях развития опухоли. Эти результаты подчеркивают значимость эволюционного аспекта в изучении рака.

Финансирование: Исследование поддержано бюджетным проектом № FWNR-2022-0006 Министерства науки и высшего образования РФ.

Integration of phylostratigraphic and phylotranscriptomic analysis in the study of differentially expressed genes in cancer tissues

Ivanov R.[1]*, Lashin S.[1, 2], Afonnikov D.[1, 2], Matushkin Yu.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *ivanovromanart@bionet.nsc.ru*

Key words: phylostratigraphy; phylotranscriptomics; differentially expressed genes; carcinoma

Motivation and Aim: Carcinomas are among the most prevalent cancer types, affecting various human organs. Understanding the evolutionary origin and its correlation with the functional role of genes associated with these diseases is important not only for understanding the nature of cancer, but also for developing new approaches to its diagnosis and treatment. In our study, we employed phylostratigraphic and phylotranscriptomic analyses to investigate the evolutionary ages of genes involved in different stages of human carcinoma development. Phylostratigraphy is a method used to determine the evolutionary origins of genes by analyzing the presence of their orthologs – genes encoding homologous proteins that diverged due to speciation – in the genomes of diverse species. This method reveals key moments in genome evolution characterized by surges in new gene numbers and identifies genes unique to specific organisms. Phylotranscriptomics, on the other hand, merges phylostratigraphic data on gene evolutionary ages with gene expression levels. This integrated approach enables the exploration of how gene age correlates with changes in gene activity across various physiological states, adaptive responses, or developmental stages of organisms or organs. Our study aims to identify genes that are differentially expressed in cancer tissues, analyze their evolutionary ages, and assess the relationship between gene age and activity during tumor development.
Methods and Algorithms: To achieve our goal, we employed a comprehensive approach that integrates differential gene expression analysis, phylostratigraphic analysis, and phylotranscriptomic analysis. For assessing gene expression levels and identifying differentially expressed genes, we utilized the DESeq2 algorithm. To determine the evolutionary origins of genes, we conducted phylostratigraphic analysis using the Orthoweb web service. This service enables the calculation of the Phylostratigraphy Age Index (PAI) by analyzing the presence of gene orthologs across different species, with a protein sequence identity threshold, calculated by Smith-Waterman algorithm, set at 0.6. Additionally, Orthoweb facilitated the estimation of the dN/dS variability index (Divergency Index, DI), comparing divergence of gene sequences with their orthologs within the Hominid family. Furthermore, we performed phylotranscriptome analysis using the myTAI tool. This analysis allowed us to explore the dynamics of gene expression in relation to their evolutionary age across various pathological stages of carcinomas, quantified in terms of the Transcriptome Age Index (TAI) and Transcriptome Divergency Index (TDI).
Results: In this study, we analyzed gene expression data from tissues affected by various carcinomas, including hepatic adenocarcinoma, renal clear cell adenocarcinoma, intestinal adenocarcinoma, breast carcinoma, prostate adenocarcinoma, thyroid carcinoma, urothelial carcinoma of the bladder, and endometrioid carcinoma of the uterine body. We also included samples from corresponding healthy tissues obtained from the TCGA database. For each cancer type, tumors were categorized into four pathological stages to assess progression. Our phylostratigraphic analysis identified two significant peaks of gene representation occurring at the early multicellular and vertebrate evolutionary stages (PAI = 3 and 6). Through phylotranscriptome analysis, we observed variations in the transcriptome age across different tumor development stages. Notably, in both early and late pathological stages of carcinomas, there was an increase in transcriptome age in cancerous tissues, whereas a decrease was noted in intermediate stages. This pattern was particularly distinct in intestinal adenocarcinomas, where the opposite trend was observed. Additionally, our findings suggest that healthy

tissues exhibit a significantly higher transcriptome age compared to their cancerous counterparts, indicating a potential upregulation of ancient genes in cancer.
Conclusion: The analysis of differential gene expression and their evolutionary age in carcinomas revealed significant differences between cancerous and healthy tissues, as well as specific changes in the dynamics of transcriptome age at different stages of tumor development. These results emphasize the importance of the evolutionary aspect in the study of cancer.
Funding: The study was supported by the Budget Project # FWNR-2022-0006 of the Ministry of Science and Higher Education of The Russian Federation.

doi 10.18699/bgrs2024-4.2-08

Прогноз свойств биологических последовательностей на основе их структурных формул – новый подход к представлению данных в биоинформатике

Лагунин А.А.[1, 2]*, Смирнов А.С.[1], Задорожный А.Д.[1], Лебедев Н.В.[1], Захаров О.С.[1], Полусмак И.В.[1], Журавлева С.И.[1], Степанян А.А.[1], Алимова А.Р.[3], Рудик А.В.[2], Филимонов Д.А.[2]

[1] *Российский национальный исследовательский медицинский университет им. Н.И. Пирогова, Москва, Россия*

[2] *НИИ биомедицинской химии им. В.Н. Ореховича, Москва, Россия*

[3] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

* *alexey.lagunin@ibmc.msk.ru*

Ключевые слова: биологические последовательности; структурные формулы; прогноз; связь структура-свойство

Мотивация и цель: Анализ структурных и функциональных свойств молекул белков, РНК и ДНК является одной из важнейших задач в молекулярной биологии. Несмотря на активное исследование в этой области, огромное количество белков, РНК и участков геномов разных организмов нуждаются в оценках этих свойств для понимания их роли в различных биологических процессах. Экспериментальные методы ограничены в своих возможностях из-за сложности, стоимости и продолжительности таких исследований. Начиная с середины XX в. для анализа структурных и функциональных свойств биологических макромолекул используют аминокислотные (а.к.) и нуклеотидные (н.т.) последовательности, описанные в виде однобуквенного кода. И это является основой большинства современных методов биоинформатики, с помощью которых можно эффективно оценивать свойства биологических макромолекул. Несмотря на значительные успехи в создании таких методов, само представление а.к. и н.т. последовательностей в виде однобуквенного кода является ограничением для дальнейшего совершенствования биоинформатических методов, так как не отражает особенностей структуры этих последовательностей на уровне связей между атомами. С другой стороны, развитие экспериментальных методов определения 3D структур биологических макромолекул позволяет описывать их атомы в трехмерном пространстве. Такое описание позволяет оценивать их характеристики и свойства намного качественней, но требует применения значительных вычислительных ресурсов, а также ограничено количеством известных трехмерных структур.

Методы и алгоритмы: Наши исследования предлагают альтернативу представления а.к. и н.т. последовательностей белков в виде структурных формул их фрагментов, что позволяет использовать разработанные ранее в хемоинформатике методы выявления связей «структура-свойство». В качестве описания структурных формул фрагментов макромолекул используются подструктурные дескрипторы многоуровневых атомных окрестностей MNA (Multilevel Neighborhoods of Atoms) высокого уровня, реализованные в специально

разработанной для работы с такими данными программе MultiPASS, использующей байесовский алгоритм для выявления связи «структура-свойство».
Результаты: Эффективность такого подхода продемонстрирована нами в методах по прогнозу: патогенных аминокислотных замен (https://www.way2drug.com/SAV-Pred/) [1, 2], аминокислотных замен, связанных с лекарственной резистентностью опухолей [3], специфичности Т-клеточных рецепторов к эпитопам (https://www.way2drug.com/TCR-Pred/) [4] и посттрансляционных модификаций [5]. В докладе будет также продемонстрировано использование нашего подхода для предсказания вторичных структур и функциональных мотивов белков, а также мишеней микроРНК и сайтов связывания транскрипционных факторов человека.
Выводы: Предложенный подход представляет новое направление биоинформатики в исследовании свойств биологических макромолекул на основе структурных формул их фрагментов.
Финансирование: Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 гг.) (№ 122030100170-5).

Prediction of properties of biological sequences based on their structural formulas – a new approach to data representation in bioinformatics

Lagunin A.A.[1, 2*], Smirnov A.S.[1], Zadorozhny A.D.[1], Lebedev N.V.[1], Zakharov O.S.[1], Polusmak I.V.[1], Zhuravleva S.I.[1], Stepanyan A.A.[1], Alimova A.R.[3], Rudik A.V.[2], Filimonov D.A.[2]

[1] *Pirogov Russian National Research Medical University (RNRMU), Moscow, Russia*
[2] *Institute of Biomedical Chemistry (IBMC), Moscow, Russia*
[3] *Lomonosov Moscow State University, Moscow, Russia*
* *alexey.lagunin@ibmc.msk.ru*

Key words: biological sequences; structural formulas; prediction; structure-property relationships

Motivation and Aim: Analysis of the structural and functional properties of protein, RNA and DNA molecules is one of the most important tasks in molecular biology. Despite active research in this area, a huge number of proteins, RNAs and regions of the genomes of different organisms require assessments of these properties to understand their role in various biological processes. Experimental methods are limited in their capabilities due to the complexity, cost, and duration of such studies. Since the mid-20th century, amino acid (aa) and nucleotide (nt) sequences, described as a single-letter code, have been used to analyze the structural and functional properties of biological macromolecules. This is the basis of most modern bioinformatics methods, which can be used to effectively evaluate the properties of biological macromolecules. Despite significant advances in the creation of such methods, the very idea of the representation of aa and nt sequences in the form of a single-letter code is a limitation for further improvement of bioinformatics methods, since it does not reflect the structural features of these sequences at the level of connections between atoms. On the other hand, the development of experimental methods for determining the 3D structures of biological macromolecules makes it possible to describe their atoms in three-dimensional space. Such a description

makes it possible to evaluate their characteristics and properties much better, but requires the use of significant computing resources, and is also limited by the number of known three-dimensional structures.

Methods and Algorithms: Our research offers an alternative to representing aa and nt sequences of proteins in the form of structural formulas of their fragments, which makes it possible to use methods previously developed in chemoinformatics for identifying structure-property relationships. To describe the structural formulas of fragments of macromolecules, high-level substructural descriptors of multilevel atomic neighborhoods MNA (Multilevel Neighborhoods of Atoms) are used, implemented in specially developed for working with such data MultiPASS program, which uses the Bayesian algorithm to identify the "structure-property" relationships.

Results: We have demonstrated the effectiveness of this approach in methods for predicting: pathogenic amino acid substitutions (https://www.way2drug.com/SAV-Pred/) [1, 2], amino acid substitutions associated with drug resistance of tumors [3], specificity of T-cell receptors for epitopes (https://www.way2drug.com/TCR-Pred/) [4] and post-translational modifications [5]. The talk will also demonstrate the use of our approach to predict secondary structures and functional motifs of proteins, as well as microRNA targets and binding sites for human transcription factors.

Conclusion: The proposed approach represents a new direction in bioinformatics in the study of the properties of biological macromolecules based on the structural formulas of their fragments.

Funding: This work has been supported within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030) (No. 122030100170-5).

Список литературы/References

1. Zadorozhny A., Smirnov A., Filimonov D., Lagunin A. Prediction of pathogenic single amino acid substitutions using molecular fragment descriptors. *Bioinformatics*. 2023;39(8):btad484. doi 10.1093/bioinformatics/btad484
2. Zadorozhny A.D., Rudik A.V., Filimonov D.A., Lagunin A.A. SAV-Pred: a freely available web application for the prediction of pathogenic amino acid substitutions for monogenic hereditary diseases studied in newborn screening. *Int J Mol Sci*. 2023;24:2463. doi 10.3390/ijms24032463
3. Zhuravleva S.I., Zadorozhny A.D., Shilov B.V., Lagunin A.A. Prediction of amino acid substitutions in abl1 protein leading to tumor drug resistance based on "Structure-Property" relationship classification models. *Life*. 2023;13:1807. doi 10.3390/life13091807
4. Smirnov A.S., Rudik A.V., Filimonov D.A., Lagunin A.A. TCR-Pred: A new web-application for prediction of epitope and MHC specificity for CDR3 TCR sequences using molecular fragment descriptors. *Immunology*. 2023;169(4):447-453. doi 10.1111/imm.13641
5. Karasev D.A., Savosina P.I., Sobolev B.N., Filimonov D.A., Lagunin A.A. Application of molecular descriptors for recognition of phosphorylation sites in amino acid sequences. *Biomed Khim*. 2017;63(5):423-427. doi 10.18097/PBMC20176305423

doi 10.18699/bgrs2024-4.2-09

Систематический анализ генетических вариантов с конфликтующими интерпретациями клинической значимости

Лазарева Т.*, Барбитов Ю.
ФГБНУ «НИИ акушерства, гинекологии и репродуктологии им. Д.О. Отта», Санкт-Петербург, Россия
** tatiana.ev.lazareva@gmail.com*

Ключевые слова: генетический вариант; интерпретация вариантов; ClinVar

Мотивация и цель: Секвенирование нового поколения (NGS) все чаще используется в биомедицинских исследованиях как мощный инструмент для выявления генетических причин наследственных заболеваний. Несмотря на разработанные рекомендации и предложенные стандартизированные классификации генетических вариантов Американского колледжа медицинской генетики (ACMG), расхождения в клинической интерпретации изменений последовательности ДНК между лабораториями сохраняются [1–3]. В настоящем исследовании на основе общедоступной базы данных ClinVar мы систематически проанализировали свойства генетических вариантов и свойства генов, для которых в базе данных о клинической значимости генетических вариантов зафиксированы расхождения в клинической интерпретации, что в конечном итоге позволит разработать стратегии для повышения согласованности интерпретации данных NGS [4].
Методы и алгоритмы: Для анализа были собраны данные ClinVar за период апрель 2018 – апрель 2024 г. о вариантах в генах, связанных с редкими заболеваниями, описанными в OMIM и представленными в Ensembl BioMart [5]. Под вариантами, для которых описан конфликт интерпретаций клинической значимости (COI), в настоящем исследовании подразумеваются варианты, которые были аннотированы как COI по крайней мере один раз за выбранный период. С помощью гипергеометрического теста были выявлены гены, обогащенные COI вариантами. Проанализированы частоты аллелей (на основе данных gnomAD genomes v2.1 [6]) и эффекты вариантов на продукт гена. Оценена эволюционная консервативность, уровень экспрессии изоформ, длина канонического транскрипта, количество экзонов, вовлеченность в биологические процессы, а также количество и тип наследования ассоциированных заболеваний, связанных с генами, для которых (не) описаны COI варианты.
Результаты: Всего для 4731 гена с описанным в OMIM наследственным заболеванием обнаружено 2 296 245 вариантов в ClinVar, в 78 % которых идентифицирован как минимум один COI вариант. Статистически значимое обогащение COI вариантов наблюдалось для 285 генов. Такие гены более сложно организованы, что следует из большего числа экзонов и большей длины транскриптов (p-value < 2.2×10^{-16}). Интересно, что эти гены имеют и большее число связанных заболеваний (p-value < 2.2×10^{-16}), преимущественно с аутосомно-доминантным

типом наследования (p-value = 3.78×10^{-9}). Выявленные гены обогащены терминами, ассоциированными со строением и функционированием мышечной ткани (например, развитие и сокращение поперечно-полосатой скелетной и сердечной мускулатуры, связывание актина и кальмодулина, функционирование трансмембранных ионных каналов).
Выводы: Полученные наблюдения согласуются с предыдущим исследованием феномена «фенотипической гетерогенности», вариабельности тяжести заболевания или даже развитием различных заболеваний у лиц, являющихся носителями патогенных вариантов в одном и том же гене [7]. Данный феномен может объяснять накопление конфликтующих интерпретаций вариантов в генах с несколькими ассоциированными заболеваниями (преимущественно из групп нервно-мышечной и сердечно-сосудистой патологии), с преобладанием аутосомно-доминантного типа наследования. Обнаруженные ассоциации между характеристиками генов и обнаружением вариантов с COI могут быть использованы для создания дополнительных правил аннотации и приоритизации вариантов в этих генах.
Финансирование: Исследование поддержано Министерством науки и высшего образования Российской Федерации (проект «Многоцентровая научно-исследовательская коллекция биоресурсов» «Репродуктивное здоровье человека» (контракт № 075-15-2021-1058 от 28 сентября 2021 г.).

Systematic analysis of genetic variants with conflicting interpretations of pathogenicity

Lazareva T.*, Barbitoff Y.
Research Institute of Obstetrics, Gynecology and Reproductology named after D.O. Ott, St. Petersburg, Russia
** tatiana.ev.lazareva@gmail.com*

Key words: genetic variant; variant interpretation; ClinVar

Motivation and Aim: Next-generation sequencing (NGS) has revolutionized biomedical research by enabling the identification of genetic causes of inherited diseases. However, discrepancies persist in the clinical interpretation of DNA variants between laboratories, despite standardized classifications proposed by the American College of Medical Genetics (ACMG) [1–3]. The study aimed to systematically analyze properties of genetic variants and their associated genes for which discordance of variant interpretation have been reported in the ClinVar database [4]. It will allow us to develop strategies for improving the consistency of NGS data interpretation.
Methods and Algorithms: We collected ClinVar data (April 2018 – April 2024) on variants in genes linked to rare diseases from OMIM (Online Mendelian Inheritance in Man) and submitted to Ensembl BioMart [5]. Variants with conflicting interpretations (COI) reported at least once during this period were defined as discordant. We employed the hypergeometric test to identify genes enriched for COI variants. Allele frequencies (gnomAD v2.1 genomes [6]) and variant effects on gene products were analyzed. Additionally, we evaluated evolutionary conservation, isoform expression levels, transcript length, exon number, involvement in biological processes, and the number and type of inheritance of diseases associated with genes harboring (or lacking) COI variants.

Results: Our analysis of 2,296,245 variants from 4,731 OMIM-linked disease genes in ClinVar revealed that 78 % contained at least one COI variant. We identified 285 genes with statistically significant enrichment of COI variants. These genes exhibited increased structural complexity, evidenced by a higher number of exons and longer transcripts (p-value $< 2.2 \times 10^{-16}$). Notably, these genes were also associated with a higher number of diseases per gene (p-value $< 2.2 \times 10^{-16}$), predominantly with autosomal dominant inheritance (p-value $= 3.78 \times 10^{-9}$). Moreover, these genes were enriched for terms related to muscle tissue structure and function.

Conclusion: Our findings align with previous studies on "phenotypic heterogeneity", where individuals with pathogenic variants in the same gene can exhibit variable disease severity or even develop different diseases [7]. This phenomenon might explain the accumulation of COI variants in genes with multiple associated diseases, particularly those related to neuromuscular and cardiovascular pathologies, often exhibiting autosomal dominant inheritance. The observed associations between gene characteristics and COI variant detection could be utilized for establishing additional rules for annotation and prioritization of variants in these genes.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (project "Multicenter research bioresource collection" "Human Reproductive Health" contract No. 075-15-2021-1058 from September 28, 2021).

Список литературы/References

1. Richards S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-423. doi 10.1038/gim.2015.30
2. Harrison S., Dolinsky J., Knight Johnson A. et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genet Med.* 2017;19:1096-1104. doi 10.1038/gim.2017.14
3. Bland A., Harrington E., Dunn K. et al. Clinically impactful differences in variant interpretation between clinicians and testing laboratories: a single-center experience. *Genet Med.* 2018;20:369-373. doi 10.1038/gim.2017.212
4. Landrum M.J. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42(D1):D980-D985. doi 10.1093/nar/gkt1113
5. Kinsella Rh. J. et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* 2011;2011:bar030. doi 10.1093/database/bar030
6. Karczewski K.J., Francioli L.C., Tiao G.et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581:434-443. doi 10.1038/s41586-020-2308-7
7. Lazareva T.E. et al. Statistical dissection of the genetic determinants of phenotypic heterogeneity in genes with multiple associated rare diseases. *Genes.* 2023;14(11):2100. doi 10.3390/genes14112100

doi 10.18699/bgrs2024-4.2-10

Ассоциация полиморфизма rs753482 гена eNOS с риском абдоминального ожирения у больных шизофренией, принимающих атипичные антипсихотики

Меднова И.А.*, Пожидаев И.В., Падерина Д.З., Петкун Д.А., Корнетова Е.Г., Иванова С.А.

НИИ психического здоровья, Томский НИМЦ РАН, Томск, Россия

** i.mednova@yandex.ru*

Ключевые слова: фармакогенетика; шизофрения; метаболические нарушения; однонуклеотидный полиморфизм

Мотивация и цель: Шизофрения представляет собой тяжелое хроническое психическое расстройство, для лечения которого необходим практически пожизненный прием антипсихотических препаратов. Наиболее часто в терапии заболевания используют антипсихотики второго поколения (атипичные антипсихотики), прием которых характеризуется развитием метаболических побочных эффектов [1]. Распространенность метаболических нарушений среди пациентов, получавших антипсихотики второго поколения, в три раза выше, чем у лиц, принимавших препараты первого поколения [2]. Одним из таких нарушений является повышение массы тела: так, на фоне приема атипичных антипсихотиков клинически значимое увеличение массы тела может наблюдаться у 50 % пациентов, при этом среди пациентов, которым впервые назначают терапию этот показатель возрастает до 70 % [3]. Ранее был показан вклад генетической компоненты [4] и эндотелиальной дисфункции в формирование метаболических нарушений при шизофрении [5]. Ген, кодирующий эндотелиальную NO-синтазу (eNOS), у человека расположен на хромосоме 7q36. Полиморфизм rs753482 eNOS приводит к появлению новой короткой стабильной формы eNOS с измененной активностью фермента, влияющей на выработку NO и клеточную функцию [6]. Однако исследований по изучению ассоциаций данного полиморфизма с метаболическими нарушениями при шизофрении не представлено. Таким образом, целью данного исследования было провести ассоциативный анализ полиморфизма rs753482 *eNOS* с абдоминальным ожирением у больных шизофренией, получающих антипсихотическую терапию.

Методы и алгоритмы: В исследование включили 485 больных шизофренией (F20 по МКБ-10), из которых 197 человек (102 женщины и 95 мужчин, средний возраст 36.42 ± 10.5 года) получали в качестве базисной терапии антипсихотики второго поколения, а 288 человек (135 женщин и 153 мужчины, средний возраст 43.31 ± 11.6 года) – антипсихотики первого поколения. Пациентам проводили антропометрическое исследование с использованием медицинского прибора Omron BF508. Наличие абдоминального ожирения устанавливали при окружности талии ≥94 см у мужчин или ≥80 см у женщин. Венозную кровь отбирали у пациентов утром натощак. ДНК получали с использованием стандартного фенол-хлороформного микрометода. Генотипирование полиморфизма осуществляли методом ПЦР в реальном времени на амплификаторе QuantStudio 5 (Applied Biosystems,

США) с использованием наборов TaqMan Assays (оборудование расположено на базе Центра коллективного пользования «Медицинская геномика» Томского НИМЦ). Статистический анализ проводился с использованием программы R версии 4.0.4. Различия считались статистически значимыми при уровне значимости $p < 0.05$.

Результаты: Абдоминальное ожирение выявляли у 101 (51.8 %) больных шизофренией, получавших в качестве базисной терапии атипичные антипсихотики и у 124 (43 %) человек, получавших лечение типичными антипсихотическими препаратами. С помощью логистической регрессии с включением пола и возраста в качестве ковариат была показана ассоциация полиморфизма rs753482 *eNOS* с абдоминальным ожирением ($p = 0.033$). По результатам ассоциативного анализа было обнаружено, что аллель А связан с повышенным риском абдоминального ожирения (OR=1.96, 95 % CI 1.12–3.41), а аллель С – со сниженным риском (OR = 0.51, 95 % CI 0.29–0.89; $p = 0.016$). Кроме того, у пациентов с генотипом АА наблюдали статистически значимое увеличение окружности талии по сравнению с показателем у пациентов с генотипами АС и СС ($p = 0.05$ и $p = 0.039$). У пациентов с генотипом АА выявили статистически значимое увеличение содержания висцерального жира по сравнению с данными, полученными у лиц с генотипом АС ($p = 0.025$). Частоты генотипов и аллелей полиморфного варианта rs753482 *eNOS* значимо не различались между пациентами с нормальной массой тела (ИМТ < 25 кг/см^2), предожирением (25 кг/см^2 ≤ ИМТ < 30 кг/см^2) и ожирением (ИМТ ≥ 30 кг/см^2). Статистически значимых различий показателей ИМТ, массы тела и содержания жира в организме у пациентов с различными генотипами, получавших лечение антипсихотиками второго поколения, не выявлено. На фоне терапии антипсихотиками первого поколения не обнаружено статистически значимых ассоциаций между исследуемым полиморфизмом и абдоминальным ожирением или показателем ИМТ. Также не выявлено статистически значимых различий показателей окружности талии, содержания висцерального жира, жира в организме, ИМТ, массы тела у пациентов с различными генотипами, получавших лечение типичными антипсихотическими препаратами.

Выводы: Выявлена ассоциация полиморфизма rs753482 гена eNOS с абдоминальным ожирением у больных шизофренией, принимающих атипичные антипсихотики в качестве базисной терапии. Представляется перспективным дальнейшее изучение молекулярно-генетических особенностей влияния антипсихотической терапии на формирование метаболических нарушений с целью реализации индивидуального подхода к терапевтической тактике.

Финансирование: Исследование поддержано грантом Российского научного фонда № 23-75-10088, https://rscf.ru/project/23-75-10088/.

Association between eNOS rs753482 polymorphism and abdominal obesity in patients with schizophrenia taking atypical antipsychotics

Mednova I.A.*, Pozhidaev I.V., Paderina D.Z., Petkun D.A., Kornetova E.G., Ivanova S.A.

Mental Health Research Institute, Tomsk National Research Medical Center, RAS, Tomsk, Russia

** i.mednova@yandex.ru*

Key words: pharmacogenetics; schizophrenia; metabolic disorders; single nucleotide polymorphism

Motivation and Aim: Schizophrenia is a severe, chronic mental disorder that requires almost lifelong treatment with antipsychotic medications. Most often, second-generation antipsychotics (atypical antipsychotics) are used in the treatment of the disease, the use of which is characterized by the development of metabolic side effects [1]. The prevalence of metabolic disorders among patients receiving second-generation antipsychotics is three times higher than in those receiving first-generation drugs [2]. One of these disorders is weight gain: while taking atypical antipsychotics, a clinically significant increase in body weight can be observed in 50 % of patients, while among patients who are prescribed therapy for the first time, this figure increases to 70 % [3]. The contribution of the genetic component [4] and endothelial dysfunction in the formation of metabolic disorders in schizophrenia [5] was previously shown. The gene encoding endothelial NO synthase (eNOS) is located on human chromosome 7q36.The rs753482 polymorphism of the eNOS gene leads to the emergence of a new short, stable form of eNOS with altered enzyme activity, affecting NO production and cellular function [6]. However, there are no studies examining the associations of this polymorphism with metabolic disorders in schizophrenia. Thus, the purpose of this study was to conduct an association analysis of the rs753482 *eNOS* polymorphism with abdominal obesity in patients with schizophrenia receiving antipsychotic therapy.
Methods and Algorithms: The study included 485 patients with schizophrenia (F20 according to ICD-10), of whom 197 persons (102 women and 95 men, average age 36.42 ± 10.5 years) received second-generation antipsychotics as basic therapy, and 288 people (135 women and 153 men, average age 43.31 ± 11.6 years) – first generation antipsychotics. Patients underwent an anthropometric study using an Omron BF508 medical device. The presence of abdominal obesity was defined as a waist circumference of ≥94 cm in men or ≥80 cm in women. Venous blood was collected from patients in the morning on an empty stomach. DNA was prepared using a standard phenol-chloroform micromethod. Genotyping of polymorphism was carried out by real-time PCR on a QuantStudio 5 amplifier (Applied Biosystems, USA) using TaqMan Assays kits (the equipment is located at the Center for Collective Use "Medical Genomics" of the Tomsk National Research Medical Center). Statistical analyzes were performed using R version 4.0.4. Differences were considered statistically significant at a significance level of $p < 0.05$.
Results: Abdominal obesity was detected in 101 (51.8 %) patients with schizophrenia who received atypical antipsychotics as basic therapy and in 124 (43 %) people treated with typical antipsychotics. Using logistic regression including sex and age as covariates, an association of the rs753482 *eNOS* polymorphism with abdominal obesity was shown ($p = 0.033$). According to the results of the association analysis, it was found that the A allele is associated with an increased risk of abdominal obesity (OR=1.96, 95 % CI 1.12–3.41), and allele C – with a reduced risk (OR=0.51, 95 % CI 0.29–0.89; $p = 0.016$). In addition, patients with the AA genotype had a statistically significant increase in waist circumference compared to patients with the AC and CC genotypes ($p = 0.05$ and $p = 0.039$). In patients with the AA genotype, a statistically significant increase in visceral fat content was revealed compared with the data obtained in individuals with the AC genotype ($p = 0.025$). The frequencies of genotypes and alleles of the polymorphic variant rs753482 *eNOS* did not differ significantly between patients with normal body weight (BMI < 25 kg/cm^2), preobesity (25 kg/cm^2 ≤ BMI < 30 kg/cm^2)

and obesity (BMI ≥ 30 kg/cm^2). There were no statistically significant differences in BMI, body weight and body fat in patients with different genotypes treated with second-generation antipsychotics. During treatment with first-generation antipsychotics, no statistically significant associations were found between the studied polymorphism and abdominal obesity or BMI. There were also no statistically significant differences in waist circumference, visceral fat, body fat, BMI, or body weight in patients with different genotypes treated with typical antipsychotic drugs.

Conclusion: An association of the rs753482 polymorphism of the eNOS gene with abdominal obesity was revealed in patients with schizophrenia taking atypical antipsychotics as basic therapy. It seems promising to further study the molecular genetic features of the influence of anipsychotic therapy on the formation of metabolic disorders in order to implement an individual approach to therapeutic tactics.

Funding: The study is supported by the Russian Science Foundation, grant number 23-75-10088. https://rscf.ru/en/project/23-75-10088/

Список литературы/References

1. Озорнин А.С., Говорин Н.В., Сахаров А.В. Содержание неэстерифицированных жирных кислот и некоторых адипокинов в сыворотке крови у пациентов с первым эпизодом шизофрении и характер их изменений при антипсихотической терапии. *Сибирский вестник психиатрии и наркологии*. 2024;1(122):51-61. doi 10.26617/1810-3111-2024-1(122)-51-61
2. De Hert M., Schreurs V., Sweers K., Van Eyck D., Hanssens L., Šinko S., van Winkel R. Typical and atypical antipsychotics differentially affect long-term incidence rates of the metabolic syndrome in first-episode patients with schizophrenia: a retrospective chart review. *Schizophrenia Res*. 2008;101(1-3):295-303. doi 10.1016/j.schres.2008.01.028
3. Bak M., Fransen A., Janssen J., van Os J, Drukker M. Almost all antipsychotics result in weight gain: a meta-analysis. *PLoS One*. 2014;9(4):e94112. doi 10.1371/journal.pone.0094112
4. Boiko A.S., Pozhidaev I.V., Paderina D.Z., Bocharova A.V., Mednova I.A., Fedorenko O.Y., Ivanova S.A. Search for possible associations of fto gene polymorphic variants with metabolic syndrome, obesity and body mass index in schizophrenia patients. *Pharmacogenomics Pers Med*. 2021;14:1123-1131. doi 10.2147/PGPM.S327353
5. Fattakhov N., Smirnova L., Atochin D. et al. Haplotype analysis of endothelial nitric oxide synthase (NOS3) genetic variants and metabolic syndrome in healthy subjects and schizophrenia patients. *Int J Obes*. 2018;42:2036-2046. doi 10.1038/s41366-018-0124-z
6. Galluccio E., Cassina L., Russo I., Gelmini F., Setola E., Rampoldi L., Casari G. A novel truncated form of eNOS associates with altered vascular function. *Cardiovasc Res.* 2014;101(3):492-502. doi 10.1093/cvr/cvt267

doi 10.18699/bgrs2024-4.2-12

Спектр CNV у пациентов с нарушениями психомоторного развития

Федотов Д.*, Кашеварова А., Скрябин Н., Лопаткина М., Васильева О., Толмачева Е., Беляева Е., Минайчева Л., Петрова В., Равжаева Е., Салюкова О., Сивоха В., Фадюшина С., Сеитова Г., Назаренко Л., Лебедев И.

Научно-исследовательский институт медицинской генетики, Томский НИМЦ РАН, Томск, Россия

** dmitry.fedotov@medgenetics.ru*

Ключевые слова: вариации числа копий участков ДНК; CNV; нарушения психомоторного развития; интеллектуальные нарушения

Мотивация и цель: Согласно МКБ-11, нарушения психомоторного развития (НПР) – группа поведенческих и когнитивных расстройств, возникающих в период развития нервной системы (интеллектуальные нарушения, расстройство развития языка или речи, расстройство аутистического спектра, расстройство развития учебных навыков, расстройство развития координации движений, синдром дефицита внимания с гиперактивностью, а также синдром стереотипных движений). К НПР могут приводить как средовые, так и генетические факторы. С развитием методов молекулярного кариотипирования был выявлен такой тип хромосомного дисбаланса, вносящий значительный вклад в этиологию НПР, как вариации числа копий участков ДНК (copy number variation, CNV), которые представляют собой фрагмент ДНК размером более одной тысячи пар нуклеотидов, по числу копий отличающийся от референсного генома [1]. Ежегодно в базе данных NCBI PubMed публикуется свыше 100 исследований относительно спектра CNV среди пациентов с НПР. В результате появляется необходимость в систематизации имеющейся информации. Данное исследование представляет собой комплексный анализ спектра CNV, опубликованного в мировой литературе, а также сравнение имеющейся информации с собственными данными.

Методы и алгоритмы: На первом этапе исследования был проведен поиск литературы в базе данных NCBI PubMed с использованием двух поисковых запросов: “neurodevelopmental disorders, copy number variation” (n = 1410) и “neurodevelopmental disorders, aCGH” (n = 208). В анализ включались работы, в рамках которых были обследованы только пациенты с НПР на предмет наличия CNV в их геномах. При этом публикации должны были содержать полный список обнаруженных аберраций с указанием версии сборки генома, все CNV должны были быть классифицированы по клинической значимости. Кроме этого, публикация должна была быть на русском или английском языке и находиться в открытом доступе. В конечный анализ включались исследования, которые содержали вышеуказанную информацию (n = 7) [2–8]. На втором этапе была сформирована собственная выборка пациентов с НПР. Пробанды были обследованы с помощью матричной сравнительной геномной гибридизации на микрочипах «SurePrint G3 Human CGH Microarray, 8×60K (Agilent Technologies, США). При доступности материала родителей проводилась количественная ПЦР в режиме реального времени с целью верификации варианта и установления его

происхождения. Все CNV классифицировали на пять групп в соответствии с рекомендациями Американской коллегии по медицинской генетике и геномике совместно с проектом «Клинический геномный ресурс» [9]. Индивиды, имеющие вероятно доброкачественные и доброкачественные аберрации, а также CNV размером более 10 Мб, исключались из исследования. Для исключения возможного эффекта взаимодействия нескольких вариантов в конечные группы были включены пациенты с одной патогенетически значимой CNV – 345 (литературные данные) и 209 (собственные данные). На последнем этапе проводилось сравнение спектров CNV относительно числа копий, происхождения и размера. Статистическая обработка данных проводилась с помощью критерия хи-квадрат. Результат считался значимым при $p < 0.05$.

Результаты: Частота патогенных CNV в собственных выборке превышает количество данных аберраций, описанных в литературе – 37.3 и 25.2 % соответственно ($p = 0.003$) (табл. 1).

Таблица 1. Характеристика CNV в сравниваемых группах

Характеристика CNV	Литературные данные	Собственные данные
Клиническая значимость		
Патогенные	87/345 (25.2 %)	78/209 (37.3 %)
Вероятно патогенные	54/345 (15.7 %)	17/209 (8.1 %)
Неопределенной клинической значимости	204/345 (59.1 %)	114/209 (54.5 %)
Число копий		
Микроделеции	162/345 (46.9 %)	97/209 (46.4 %)
Микродупликации	182/345 (52.7 %)	109/209 (52.2 %)
Микротрипликации	1/345 (0.3 %)	3/209 (1.4 %)
Происхождение		
De novo	55/150 (36.7 %)	15/64 (23.4 %)
Материнское	55/150 (36.7 %)	25/64 (39.1 %)
Отцовское	40/150 (26.7 %)	24/64 (37.5 %)
Размер		
До 0.5 м.п.о.	175/345 (50.7 %)	145/209 (69.4 %)
0.5–1 м.п.о.	48/345 (13.9 %)	21/209 (10.0 %)
1–5 м.п.о.	106/345 (30.7 %)	36/209 (17.2 %)
5–10 м.п.о.	16/345 (4.6 %)	7/209 (3.3 %)

При этом значительное количество вариантов было классифицировано как CNV с неопределенной клинической значимостью в обеих группах, что связано с трудностями клинической интерпретации не рекуррентных вновь выявляемых аберраций. В выборке CNV как из литературных, так и собственных данных микроделеции и микродупликации встречались с сопоставимой частотой, с незначительным преобладанием микродупликаций – 52.7 и 52.2 %. Микродупликации традиционно считаются менее патогенными, зачастую имея менее выраженный клинический эффект, и могут расцениваться как варианты с неопределенной клинической значимостью. В обеих группах чаще встречаются CNV, унаследованные от родителей, по сравнению с возникшими *de novo* – 63 % (литературные данные) и 76 % (собственные данные) ($p = 0.059$). Унаследованный характер аберрации, с одной стороны, уменьшает вероятность ее патогенности; а

с другой стороны, может указывать на ее неполную пенетрантность и наличие дополнительных генетических, эпигенетических и средовых факторов, влияющих на развитие заболевания. В группе литературных данных преимущественно встречаются CNV размером до 0.5 м.п.о. – 50.7 %, при этом среди CNV из собственных данных частота таких аберраций значительно выше – 69.4 % ($p < 0.001$). Размер выявляемых вариантов обусловлен, прежде всего, разрешающей возможностью метода анализа. Очевидно, что накопление крупного массива данных относительно вероятно патогенных CNV и вариантов с неопределенной клинической значимостью в одних и тех же регионах может способствовать описанию новых синдромов. В литературных и собственных данных чаще встречались CNV в регионах 7q33, 7q36.2, 8p23.3, 11q23.3, 15q11.2, 16p13.11, 16p12.3, Xp22.31, которые были классифицированы как вероятно патогенные или варианты с неопределенной клинической значимостью.
Заключение: Проведенный анализ позволил описать структуру CNV у пациентов с НПР, выделить частые варианты в одних и тех же регионах, что необходимо для более глубокого понимания их возможного патогенного эффекта. Однако высокая частота определенной CNV не может являться основным фактором для интерпретации ее как патогенного варианта. Существует необходимость в экспериментальном подтверждении патологических эффектов часто встречающихся вероятно патогенных вариантов и CNV с неопределенной клинической значимостью на модельных системах.
Финансирование: Исследование поддержано грантом Российского научного фонда № 21-65-00017, https://rscf.ru/project/21-65-00017/.

CNV spectrum in patients with neurodevelopmental disorders

Fedotov D.*, Kashevarova A., Skryabin N., Lopatkina M., Vasilyeva O., Tolmacheva E., Belyaeva E., Minaycheva L., Petrova V., Ravzhaeva E., Salyukova O., Sivoha V., Fadyushina S., Seitova G., Nazarenko L., Lebedev I.
Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia
* *dmitry.fedotov@medgenetics.ru*

Key words: copy number variation; CNV; neurodevelopmental disorders; intellectual disabilities

Motivation and Aim: According to ICD-11, neurodevelopmental disorders (NDDs) are a group of behavioral and cognitive disorders that occur during the nervous system development (intellectual disability, language or speech developmental disorder, autism spectrum disorder, learning disabilities, movement coordination disorder, attention deficit hyperactivity disorder, and stereotypic movement disorder). Both environmental and genetic factors can lead to NDDs. Due to the development of molecular karyotyping methods, a type of chromosomal imbalance that contributes significantly to the etiology of NDDs has been identified – a copy number variation (CNV), which is a DNA fragment of more than one thousand nucleotide pairs that differs in copies from the reference genome [1]. Annually, more than 100 studies regarding the spectrum of CNVs among patients with NDDs are published in the NCBI PubMed database. As a result, there is a need to systematisation of available information. This study was aimed as a comprehensive analysis of the CNV spectrum published in the world literature, and comparison of the available information with our own data.

Methods and Algorithms: In the first stage, a literature search was performed in the NCBI PubMed database using two search terms 'neurodevelopmental disorders, copy number variation' (n=1410) and 'neurodevelopmental disorders, aCGH' (n=208). Studies that examined only patients with NDDs for the presence of CNVs in their genomes were included in the analyses. In this case, publications had to contain a complete list of detected aberrations with the genome assembly version, and all CNVs had to be interpreted by clinical significance. In addition, the publication had to be in Russian or English and open access. Studies that contained the above information (n=7) were included in the final analysis [2–8]. At the second stage, our own sample of patients with NDDs was formed. The probands were screened using array comparative genomic hybridization (aCGH) on SurePrint G3 Human CGH Microarray, 8×60K (Agilent Technologies, USA). In case of parental material availability, quantitative real-time PCR was performed to verify the variant and establish its origin. All CNVs were classified into five groups according to the recommendations of the American College of Medical Genetics and Genomics in conjunction with the Clinical Genomic Resource Project [9]. Individuals with likely benign and benign aberrations and CNVs larger than 10 Mb were excluded from this study. In order to exclude the possible interaction effect of several variants, patients with one pathogenically significant CNV were included in the final groups, 345 (literature data) and 209 (own data). In the last step, set of CNV was compared with respect to the number of copies, origin and size. Statistical analysis of the data was performed using chi-square criteria. The result was considered significant at $p < 0.05$.

Results: The frequency of pathogenic CNVs in our own data exceeds the number of these aberrations found in the literature, 37.3 and 25.2 %, respectively ($p = 0.003$) (Table 1).

Table 1. Characterisation of CNVs in the compared groups

Characteristic	Literature data	Own data
Clinical significance		
Pathogenic	87/345 (25.2 %)	78/209 (37.3 %)
Likely pathogenic	54/345 (15.7 %)	17/209 (8.1 %)
Uncertain clinical significance	204/345 (59.1 %)	114/209 (54.5 %)
Number of copies		
Microdeletions	162/345 (46.9 %)	97/209 (46.4 %)
Microduplications	182/345 (52.7 %)	109/209 (52.2 %)
Microtriplications	1/345 (0.3 %)	3/209 (1.4 %)
Origin		
De novo	55/150 (36.7 %)	15/64 (23.4 %)
Maternal	55/150 (36.7 %)	25/64 (39.1 %)
Paternal	40/150 (26.7 %)	24/64 (37.5 %)
Size		
Up to 0.5 Mb	175/345 (50.7 %)	145/209 (69.4 %)
0.5–1 Mb	48/345 (13.9 %)	21/209 (10.0 %)
1–5 Mb	106/345 (30.7 %)	36/209 (17.2 %)
5–10 Mb	16/345 (4.6 %)	7/209 (3.3 %)

At the same time, a significant number of variants were classified as CNVs with uncertain clinical significance in both groups, which is associated with difficulties in clinical interpretation of non-recurrent new aberrations. In both the CNV sample from the literature and our own data microdeletions and microduplications occurred with

equal frequency with a slight predominance of microduplications, 52.7 and 52.2 %. Microduplications have traditionally been considered less pathogenic, often having less clinical effect, and may be considered as variants of uncertain clinical significance. Parentally inherited CNVs were more common in both groups, 63 % (literature data) and 76 % (own data) ($p = 0.059$). The inherited nature of the aberration, on the one hand, reduces the likelihood of its pathogenicity; on the other hand, it may indicate its incomplete penetrance and the existence of additional genetic, epigenetic and environmental factors influencing the development of the disease. CNVs as small as 0.5 Mb are predominantly found in the literature data group – 50.7 %, while the frequency of such aberrations is significantly higher among CNVs from our own data – 69.4 % ($p < 0.001$). The size of the identified variants is primarily determined by the resolution of the analysis method. Obviously, the accumulation of a large amount of data regarding likely pathogenic CNVs and variants with uncertain clinical significance in the same regions may contribute to the description of new syndromes. In the literature and our own data, CNVs in regions 7q33, 7q36.2, 8p23.3, 11q23.3, 15q11.2, 16p13.11, 16p12.3, Xp22.31 were more frequently observed and classified as likely pathogenic variants or aberrations with uncertain clinical significance.

Conclusion: This analysis made it possible to describe the structure of CNV in patients with NDD, allowed to identify frequent CNVs in the same regions, which is necessary to better understand the possible pathogenic effect of some variants. However, the high frequency of occurrence of a particular CNV cannot be the main factor for interpreting it as a pathogenic aberration. There is a need to experimentally validate the pathological effects of frequently occurring likely pathogenic variants and CNVs of uncertain clinical significance on model systems.

Funding: This study is supported by the Russian Science Foundation grant No. 21-65-00017, https://rscf.ru/project/21-65-00017/

Список литературы/References

1. Kearney H.M., Thorland E.C., Brown K.K. et al. Working group of the american college of medical genetics laboratory quality assurance committee. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med.* 2011;13(7):680-685. doi 10.1097/GIM.0b013e3182217a3a
2. Chehbani F., Tomaiuolo P., Picinelli C. et al. Yield of array-CGH analysis in Tunisian children with autism spectrum disorder. *Mol Genet Genomic Med*. 2022;10(8):e1939. doi 10.1002/mgg3.1939
3. Baccarin M., Picinelli C., Tomaiuolo P. et al. Appropriateness of array-CGH in the ADHD clinics: A comparative study. *Genes Brain Behav*. 2020;19(6):e12651. doi 10.1111/gbb.12651
4. Zarrei M., Burton C.L., Engchuan W. et al. A large data resource of genomic copy number variation across neurodevelopmental disorders. *NPJ Genom Med*. 2019;(4):26. doi 10.1038/s41525-019-0098-3
5. Lindstrand A., Eisfeldt J., Pettersson M. et al. From cytogenetics to cytogenomics: whole-genome sequencing as a first-line test comprehensively captures the diverse spectrum of disease-causing genetic variation underlying intellectual disability. *Genome Med*. 2019;11(1):68. doi 10.1186/s13073-019-0675-1
6. Lopes F., Torres F., Soares G. et al. Genomic imbalances defining novel intellectual disability associated loci. *Orphanet J Rare Dis.* 2019;14(1):164. doi 10.1186/s13023-019-1135-0
7. Firouzabadi S.G., Kariminejad R., Vameghi R. et al. Copy number variants in patients with autism and additional clinical features: Report of VIPR2 duplication and a novel microduplication syndrome. *Mol Neurobiol*. 2017;54(9):7019-7027. doi 10.1007/s12035-016-0202-y
8. Fry A.E., Rees E., Thompson R. et al. Pathogenic copy number variants and SCN1A mutations in patients with intellectual disability and childhood-onset epilepsy. *BMC Med Genet*. 2016;17(1):34. doi 10.1186/s12881-016-0294-2
9. Riggs E., Andersen E., Cherry A. et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med.* 2022;22(9):245-257. doi 10.1038/s41436-019-0686-8

doi 10.18699/bgrs2024-4.2-13

Association of pubertal trajectories with molecular markers and health outcomes in young men: prospective cohort study

Bezuglov V.[1, 2], Shtratnikova V.[1], Suvorov A.[3], Hauser R.[4], Sergeyev O.[1, 5]*

[1] *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

[2] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

[3] *Department of Environmental Health Sciences, University of Massachusetts, Amherst, MA, USA*

[4] *Environmental and Occupational Medicine and Epidemiology Program, Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

[5] *Research Centre for Medical Genetics, Moscow, Russia*

* *olegsergeyev1@yandex.ru*

Key words: puberty; pubertal progression; epigenetic program;small RNA; microRNA; piRNA; tRNA-derived small RNA; RNA-seq; sperm; hormones; semen quality

Motivation and Aim: Puberty is a hormone-dependent period in which spermatogenesis begins in the testicles. The age of pubertal onset, the pubertal progression and the age of sexual maturation varies in healthy children/adolescents. There is evidence that spermatogenesis is accompanied by dramatic changes in the epigenetic marks of sperm [1]. Particularly, in the profile of small non-coding RNAs that are less than 200 nucleotides and play a role in RNA silencing and gene expression [2]. The aim of our study was to prospectively investigate the relationship of pubertal trajectories with the reproductive outcomes, semen quality, serum hormone level and sperm sncRNA profile in healthy young adults. The secondary aim was to investigate the sperm sncRNA profile by semen quality and serum hormones.

Methods and Algorithms: Physical examinations including measurement of testicular volume (TV) by Prader orchidometer were performed at enrollment of the prospective cohort Russian Children Study (ages 8–9 years), and annually up to 19 years of age. We used group-based trajectory models (GBTMs) to identify subgroups of boys who followed similar pubertal trajectories from ages 8 till 19 years based on annual TV [3]. At 18–19 years semen samples were collected from 225 young adults. Semen was fractionated using 90 % and 45 % density gradients to isolate spermatozoa by density and maturity and to remove somatic cell contamination. Basic semen quality parameters, including semen volume, sperm concentration, motility and total progressive motile sperm count (TPMSC) were analyzed in accordance with the updated criteria by the Nordic Association for Andrology (NAFA) and European Society of Human Reproduction and Embryology–Special Interest Group in Andrology (ESHRE-SIGA) [5]. Hormone levels were analyzed using the Architect i1000SR and chemiluminescent microparticle immunoassay (Abbott Laboratories). Kruskal-Wallis rank test was used for comparison of different factors between pubertal subgroups/trajectories.

Fifty one subjects were randomly selected for analysis of molecular markers. RNA isolation from sperm was performed as described earlier [6]. RNA was purified using the RNeasy Mini Kit (Qiagen). Libraries of small RNA were constructed using NEBNext Multiplex Small RNA Library Prep Set (New England BioLabs) per manufacturer's guidelines followed by gel-size selection of 145–165 nucleotides

fragments. Libraries were sequenced on NextSeq 500 (Illumina) using one multiplexed single-end run for 75 nt.

We used the previously constructed pipeline for small RNA seq data analysis [8]. It is based on mapping reads on the reference genome and assigning them to RNA species using ITAS – integrated annotation of microRNA, piRNA and mature tRNA [7]. Reads attributed to mature tRNA were assigned to different tRNA fragments (tsRNA) using the pseudo-alignment method. Expression data matrices were filtered and differential expression analysis by DESeq2 was used separately for less mature (n=20) and more mature (n=27) sperm.

Results: Three different pubertal trajectories were identified among 225 subjects: slower, 37 %, moderate, 41 %, and faster, 18 % pubertal progression (3 % missing). Further, groups with faster and moderate trajectories were joined in one group. For slower trajectory group the birth weight, BMI and TV at semen sampling, sperm concentration, TPMSC were significantly lower and serum FSH level was significantly higher compared with faster combined with moderate one (Table 1).

Table 1. Significant different reproductive outcomes, serum hormone, and anthropometric parameters by slower vs faster pubertal trajectories

Parameters\ Trajectory	Slower				Moderate and Faster				p-value
	N	**median**	p25	p75	N	**median**	p25	p75	
Mean testicular volume, ml	77	**21.3**	20.0	23.8	116	**28.8**	26.3	31.3	0.0001
Sperm concentration, mill/ml	84	**39.9**	21.7	70.6	132	**62.7**	37.5	106.4	0.0001
TPMSC, mill	84	**55.7**	19.6	93.4	132	**88.1**	42.0	152.8	0.0002
FSH, U/l	84	**3.4**	2.4	5.1	133	**2.6**	1.7	3.4	0.0001
BMI, kg/m^2	84	**19.9**	18.6	22.3	133	**21.8**	19.7	23.6	0.0001
Birth weight, kg	82	**3.3**	3.0	3.5	132	**3.5**	3.2	3.8	0.0001

Sequencing was completed with median (25–75 percentile) 7.9 (5.1-10.7) million reads per sample and 63.56 (48.8-73.54) % of alignment rate. 153 miRNA, 362 piRNA and 59 tsRNA (a total of 574 sncRNA) were identified with strict criteria of mean count > 5 and no zero value of counts in any samples. We found 11 differentially expressed (DE) sncRNA with pubertal trajectories (1 microRNA, 3 piRNA, 7 tsRNA). Additionally, 152 sncRNA (41 microRNA, 94 piRNA, 11 tsRNA) were DE with semen quality parameters. Of these, 31 sncRNA were also DE with serum hormone levels. There were 5 microRNA, 5 piRNA and 12 tsRNA differentially expressed with more than one factor (Fig. 1). Particularly, hsa-piR-32951 was DE with pubertal trajectories, TV, sperm concentration and TPMSC. miR-93-5p was DE with semen quality (TV, TPMSC) and serum hormone levels (testosterone, LH, TSH, estradiol).

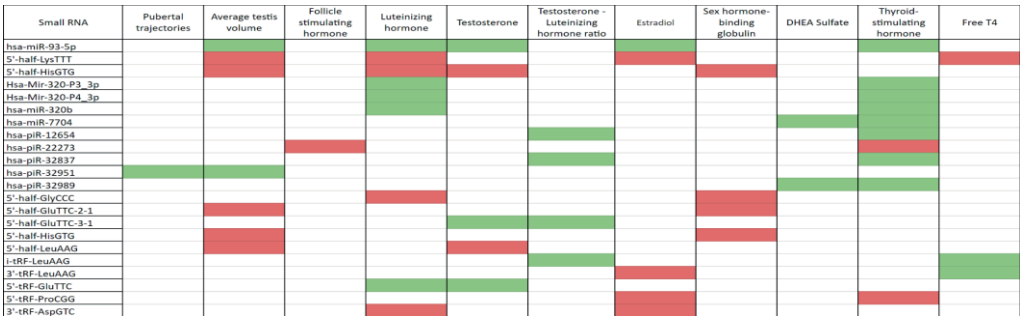


Fig. 1. Differentially expressed small RNA in less mature sperm (green) and more mature sperm (red) by pubertal trajectories, serum hormones and reproductive outcomes

Conclusion: The slower pubertal trajectory was associated with lower birth weight, lower BMI and TV at semen collection, lower TPMSC and sperm concentration and higher serum FSH. We found some sperm sncRNA as promising candidate biomarkers of pubertal trajectories and reproductive outcomes in young adulthood.

Funding: This work was funded by the Russian Science Foundation (RSF), #18-15-00202; https://rscf.ru/en/project/18-15-00202/ – sncRNA profiling; and the National Institute of Environmental Health Sciences (NIEHS), USA, #R01 ES014370 – parent prospective cohort RCS, semen and hormone analysis. We would like to thank the study participants, staff and researchers of Russian Children's Study. We also thank Maria Logacheva for her assistance in the RNA sequencing.

References

1. Nicholls P.K. et al. Hormonal regulation of sertoli cell micro-RNAs at spermiation. *Endocrinology*. 2011;152(4):1670-1683. doi 10.1210/en.2010-1341
2. Santiago J. et al. All you need to know about sperm RNAs. *Hum Reprod Update*. 2021;28(1):67-91. doi 10.1093/humupd/dmab034
3. Burns J.S. et al. Longitudinal association of prepubertal urinary phthalate metabolite concentrations with pubertal progression among a cohort of boys. *Environ Res*. 2023;233:116330. doi 10.1016/j.envres.2023.116330
4. Björndahl L. et al. A practical guide to basic laboratory andrology. Cambridge University Press, 2010
5. Shtratnikova V. et al. Optimization of small RNA extraction and comparative study of NGS library preparation from low count sperm samples. *Syst Biol Reprod Med*. 2021;67(3):230-243. doi 10.1080/19396368.2021.1912851
6. Bezuglov V. et al. Approaches for sRNA analysis of human RNA-Seq data: comparison, benchmarking. *Int J Mol Sci*. 2023;24(4):4195. doi 10.3390/ijms24044195
7. Stupnikov A. et al. ITAS: integrated transcript annotation for small RNA. *Non-coding RNA*. 2022;8(3):30. doi 10.3390/ncrna8030030

doi 10.18699/bgrs2024-4.2-14

Genes encoding Hero proteins and ischemic stroke risk: a comprehensive molecular-genetic and bioinformatics analysis

Bushueva O.

Laboratory of Genomic Research, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russia
Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, Kursk, Russia
olga.bushueva@inbox.ru

Key words: ischemic stroke; Hero proteins; SNPs; gene expression

Motivation and Aim: In 2020, Japanese researchers identified a novel class of proteins, termed "heat-resistant obscure" (Hero), known for retaining functionality even under high temperatures [1]. Hero proteins exhibit chaperone-like properties, suggesting their potentially significant role in the regulation of vascular homeostasis and neuroprotection. This study aims to comprehensively analyze 115 polymorphic variants, encompassing Hero genes (*C9orf16*, *C11orf58*, *BEX3*, *SERBP1*, *SERF2*, *C19orf53* – totaling 33 loci), chaperones, co-chaperones, and their regulators (59 loci), as well as variants identified through genome-wide association studies (GWAS) (23 loci), to investigate their correlation with the risk of developing ischemic stroke (IS).
Methods and Algorithms: Genotyping was performed on a sample of the Central Russian population comprising 2000 individuals, including 1000 IS patients and 1000 control subjects, utilizing the MassARRAY-4 genetic analyzer (USA) and Real-Time CFX96 (USA). Peripheral blood expression of Hero genes was assessed via quantitative PCR using a CFX96 instrument.
Results: Analysis revealed associations between seven Hero gene SNPs (rs10766342 *C11orf58*, rs11024032 *C11orf58*, rs11826990 *C11orf58*, rs3203295 *C11orf58*, rs10832676 *C11orf58*, rs4757429 *C11orf58*, rs12566098 *SERBP1*), two heat shock protein SNPs located in *BAG2*, *HSPA1L* genes, and six GWAS-significant SNPs located in *PITX2*, *CASZ1*, *AC016251.1*, *ATXN2*, *TWIST1* genes and ischemic stroke risk. Furthermore, gender and environmental factors such as smoking and fresh vegetable/fruit consumption were identified as modifiers of the associations between polymorphic loci of Hero genes, other chaperones, and IS risk. Analysis of intergenic (G×G) interactions using the MB-MDR method revealed that the best G×G models primarily involved polymorphic variants of the Hero genes (especially *C19orf53* and *C11orf58*) interacting with GWAS loci. Additionally, the best models of gene-environment (G×E) interactions were shaped by smoking and polymorphic loci of chaperone genes, including heat shock proteins and their regulators, alongside Hero genes. Bioinformatics analysis revealed the high regulatory potential of Hero SNPs, characterized by their influence on binding to quantitative trait loci and transcription factors. Furthermore, comparative expression analysis of Hero genes in blood samples exhibited differences between controls and acute stroke patients, particularly in the expression levels of *C19orf53* ($P = 0.01$) and *SERBP1* ($P = 0.01$).
Conclusion: This study is the first in the world to discover the role of Hero genes in the risk of IS.
Funding: The study is supported by Russian Science Foundation (No. 22-15-00288, https://rscf.ru/project/22-15-00288/).

References

1. Tsuboyama K. et al. *PLoS Biol*. 2020;18(3):e3000632. doi 10.1371/journal.pbio.3000632

doi 10.18699/bgrs2024-4.2-15

HSP90AA1 gene polymorphisms and the risk of ischemic stroke

Kobzeva K.[1]*, Dorofeeva A.[1], Bushueva O.[1, 2]

[1] *Laboratory of Genomic Research, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russia*

[2] *Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, Kursk, Russia*

* *kseniya.kobzeva.0246@gmail.com*

Key words: chaperones; SNP; ischemic stroke; HSP90AA1

Motivation and Aim: Ischemic stroke (IS), ranking as the second leading cause of mortality and the primary contributor to disability worldwide, poses a significant health burden. Heat shock proteins (HSPs) play a pivotal role in IS pathogenesis and subsequent post-stroke conditions. Notably, HSP90 emerges as a key player, influencing inflammation, endothelial dysfunction, platelet aggregation, and atherosclerosis [1–3]. Moreover, HSP90 serves crucial functions in proteome protection during stress, facilitation of polypeptide folding and transport, promotion of proteolysis, and regulation of protein complex formation [4–7]. Acting as a vital molecular chaperone, HSP90 orchestrates protein folding, intracellular disposition, and assembly of various proteins, including pivotal mediators of signal transduction and transcriptional regulation.

Methods and Algorithms: The objective of this research was to investigate the correlation between polymorphisms within gene encoding HSP90 family member, specifically *HSP90AA1*, and the susceptibility to and clinical characteristics of IS. We conducted genotyping of DNA samples obtained from 1018 IS patients and 1019 healthy controls using the MassArray-4 system for rs11621560 and rs4264324 *HSP90AA1*. Additionally, genotyping of rs7155973 *HSP90AA1* was performed via RT-PCR utilizing TaqMan probes. The assessment of associations was carried out using a log-additive regression model. Furthermore, bioinformatics tools were employed to analyze the functional implications of the identified SNPs.

Results: We identified that the SNP rs11621560 *HSP90AA1* is associated with an increased risk of IS among individuals with low intake of fresh fruits and vegetables (risk allele C, OR = 2.56, 95 % CI 1.52–4.34, P = 0.0004, Pbonf = 0.0008). Through subsequent bioinformatic analysis, we established the molecular mechanisms underlying this SNP's involvement in IS. Firstly, via cis-eQTL effects, rs11621560 *HSP90AA1* upregulates the expression of *TRAF3* (neuron-specific tumor necrosis factor receptor–associated factor 3) in the brain cortex. *TRAF3* has been implicated as a central regulator of neuronal death in acute IS, with overexpression exacerbating neuronal loss and infarct size [8]. Secondly, examination of the transcription factors (TF) binding to the DNA region created by the risk allele C rs11621560 *HSP90AA1*, along with TF-associated overrepresented biological processes, unveiled the SNP's key functions in IS, including the regulation of apoptosis (GO:0006915, GO:0043525, GO:2001224), response to stress (GO:0033554, GO:0140467) and hypoxia (GO:007145), immune modulation (GO:0071560, GO:0001819), gliogenesis (GO:0014015), neuron differentiation (GO:0045666), and blood vessel development (GO:0001568). These processes not only contribute to the onset of IS but also influence its outcomes. Additionally, we observed

an association between the allele A of rs7155973 *HSP90AA1* and international normalized ratio (INR) (Difference 0.08, 95 % CI 0.02–0.15, P = 0.0093).
Conclusion: Thus, the *HSP90AA1* gene polymorphisms represents a novel genetic risk marker for IS: specifically, rs11621560 *HSP90AA1* heightens IS risk, with fresh fruit and vegetable intake modulating this risk, while rs7155973 *HSP90AA1* impacts the coagulation parameter – INR.
Funding: The study is supported by Russian Science Foundation (No. 22-15-00288).

References

1. Wang J., Cui S., Zhang X., Wu Y., Tang H. High expression of heat shock protein 90 is associated with tumor aggressiveness and poor prognosis in patients with advanced gastric cancer. *PLoS One.* 2013;8(4):e62876. doi 10.1371/journal.pone.0062876
2. Xu X., Hua Y., Zhang H., Wu J., Miao Y. et al. Greater stress protein expression enhanced by combined prostaglandin A1 and lithium in a rat model of focal ischemia. *Acta Pharmacol Sin.* 2007;28(8):1097-1104. doi 10.1111/j.1745-7254.2007.00624.x
3. Mu H., Wang L., Zhao L. HSP90 inhibition suppresses inflammatory response and reduces carotid atherosclerotic plaque formation in ApoE mice. *Cardiovasc Ther.* 2017;35(2). doi 10.1111/1755-5922.12243
4. Terasawa K., Minami M., Minami Y. Constantly updated knowledge of Hsp90. *J Biochem.* 2005;137(4):443-447. doi 10.1093/jb/mvi056
5. Sreedhar A.S., Kalmár E., Csermely P., Shen Y.-F. Hsp90 isoforms: functions, expression and clinical importance. *FEBS Lett.* 2004;562(1-3):11-15. doi 10.1016/s0014-5793(04)00229-7
6. Liao D.F., Jin Z.G., Baas A.S., Daum G., Gygi S.P. et al. Purification and identification of secreted oxidative stress-induced factors from vascular smooth muscle cells. *J Biol Chem.* 2000;275(1):189-896. doi 10.1074/jbc.275.1.189
7. Li J., Soroka J., Buchner J. The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochim Biophys Acta.* 2012;1823(3):624-635. doi 10.1016/j.bbamcr.2011.09.003
8. Gong J., Li Z-Z., Guo S., Zhang X-J., Zhang P., Zhao G.-N. et al. Neuron-specific tumor necrosis factor receptor–associated factor 3 is a central regulator of neuronal death in acute ischemic stroke. *Hypertension.* 2015;66(3):604-616. doi 10.1161/HYPERTENSIONAHA.115.05430

doi 10.18699/bgrs2024-4.2-16

Discovery of somatic mutations associated with clonally expanded T cells

Nikonova E.O.[1, 2]*, Komech E.A.[2, 3], Barinova A.A.[2], Samitova A.F.[3], Belova V.A.[3], Buyanova A.A.[3], Repinskaya Z.A.[3], Suchalko O.N.[3], Korostin D.O.[3], Lukyanov S.[2, 4], Zvyagin I.V.[2, 3#]

[1] *Skolkovo Institute of Science and Technology, Moscow, Russia*

[2] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia*

[3] *Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russia*

[4] *Pirogov Russian National Research Medical University, Moscow, Russia*

** sakhliza@yandex.ru, # izvyagin@gmail.com*

Key words: somatic mutations; clonally expanded T cell; ankylosing spondylitis; cytomegalovirus; autoimmune diseases; T cell regulation.

Motivation and Aim: Somatic mutations may occur during cell division, and intensively dividing cells are more susceptible to acquiring such mutations. Somatic mutagenesis has been actively studied in cancer, where mutations can act as oncogenic drivers, but their role in other diseases is not fully clear [1]. We hypothesized that somatic mutations can arise in T cells proliferating after activation during immune response, and some of such mutations may have functional consequences affecting cell survival, proliferation, and thus promoting persistence and/or dysregulation of T cell clones. Recent studies reported presence of somatic mutations in different subsets of T cells from patients with autoimmune diseases [2, 3]. However, these mutations have been detected in bulk CD4+ and CD8+ T cells, and a direct association of these mutations with specific clones has not been shown. Thereby, our research aims to investigate the presence of somatic mutations in persistent clonally expanded T cells associated with different types of immunity: autoreactive and antiviral response.

Methods and Algorithms: The experimental design included T cells isolated on the basis of their antigen-specificity from peripheral blood of 2 donors (D1, D2) with chronic viral infection – cytomegalovirus (CMV) and 2 donors (D3, D4) with autoimmune disease – ankylosing spondylitis (AS). Fractions of antigen-specific CD8+ T cells for AS (HLA-B*27:05-YeiH epitope) and CMV (HLA-A*02:01-NLV epitope), as well as control T cell samples were isolated from PBMC of each donor in multiple replicates using FACS and peptide-MHC multimer staining [4]. For each cell sample (60-200 cells per sample), T cell receptor (TCR) sequencing was performed followed by clonotype enrichment analysis using MiXCR software [5, 6]. Multimer-positive samples containing at least 50% of one AS/CMV-associated T cell clone were then used for whole genome amplification, followed by whole exome sequencing (WES) averaging ~100x read coverage per sample. Somatic mutations were identified from the WES data using Mutect2 and VarScan [7]. For further analysis, we considered variants detected by Mutect2 in both replicates.

Results: YeiH- and NLV-specific T cell fractions (AS-related and CMV pp65 protein epitopes, respectively) had 1 or 2 large T cell clones comprising at least 70% of all T cells in the sample. At the same time, none of these clones were present in the cell

fractions containing multimer-negative CD8+ T cells. The obtained NLV-specific clonotypes were highly homologous with CMV-specific clonotypes from VDJdb database [8]. And YeiH-specific clonotypes matched TCR beta chain motif previously identified in other researches [9].

For each cell sample (60-200 T cells per sample) we obtained at least $90*10^6$ paired reads, of which 99% were mapped to the human reference genome (hg38). Overall read coverage reached 111 and 116 in exonic regions on average for each sample.

For D1 we found 32 somatic mutations reproduced in both replicates and confirmed by Mutect2, for D2 – 13, for D3 – 48, and for D4 – 18. It is noteworthy that the mutation frequency roughly corresponds to the clone representation frequency in the repertoire. In the NLV+ sample from D1, there are two clusters of variants reproduced in both replicates and verified by Mutect2, with a fraction of reads with mutation around 0.4 and 0.1 (Fig. 1). This corresponds to the presence of two large clones in this sample, accounting for 80% and 20% based on the TCR sequencing results.

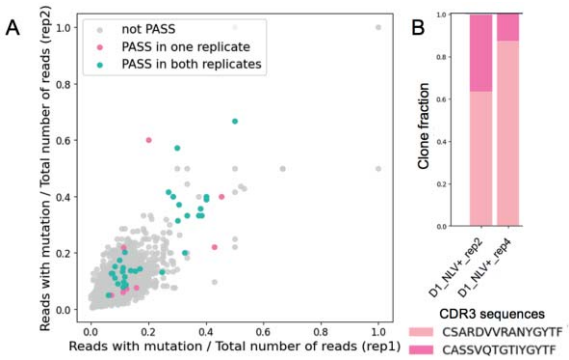


Fig. 1. A – Fraction of reads with a variant in each replicate for NLV+ samples of D1; B – Fractions of two large T cell clones based on the TCR sequencing results. PASS – variants verified by Mutect2; rep – replicate

Several mutations were found in genes involved in the regulation of cell proliferation, apoptosis, tumorigenesis, and tumor suppression. None of the detected mutations were shared at the gene level across different donors and between AS and CMV. However, some genes were grouped into common signaling pathways: MAPK, Ras, Rap1, PI3K-Akt signaling pathways; or had similar functions (Table 1).

Conclusion: Using peptide-MHC multimer cell sorting followed by whole exome sequencing of samples enriched for epitope-specific T cells, we discovered genomic variants associated with clonally expanded T cells involved in antiviral response or autoimmunity. The data obtained suggest that these variants represent somatic mutations that arise either during maturation or after activation and subsequent proliferation of naive precursors of the T cell clones. Despite the uniqueness of mutations between different donors and T cell types, some of the variants detected in different donors affect genes involved in processes of immune regulation and cell survival, implying their impact on the functional activity and/or persistence of the T cells.

Table 1. Identified somatic mutations for NLV- and YeiH-specific samples from WES data

Samples	Gene	Gene function	KEGG pathways*
NLV+	EPHA2	Involved in tumorigenesis and malignant progression	MAPK/Ras/Rap1/PI3K-Akt signaling pathways
	IGF1R	Regulates proliferation and apoptosis	MAPK/Ras/Rap1/FoxO/PI3K-Akt/AMPK/mTOR signaling pathways, Pathways in cancer, Transcriptional misregulation in cancer
YeiH+	CEBPB	Tumor suppressor, inhibits T cell proliferation	TNF/IL-17 signaling pathways, Transcriptional misregulation in cancer
	SLC8A2	Tumor suppressor	Calcium/cGMP-PKG signaling pathways
	AFDN	Inhibits cell proliferation	Ras/Rap1/cAMP signaling pathways
	DLG2	Knockout promotes cell proliferation	Hippo signaling pathway

*according to KEGG pathways database [10].

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (grant #075-15-2019-1789).

References

1. Van Horebeek L., Dubois B., Goris A. Somatic Variants: New Kids on the Block in Human Immunogenetics. *Trends Genet*. 2019;35(12):935-947. doi 10.1016/j.tig.2019.09.005
2. Valori M., Jansson L., Tienari P.J. CD8+ cell somatic mutations in multiple sclerosis patients and controls-Enrichment of mutations in STAT3 and other genes implicated in hematological malignancies. *PLoS One*. 2021;16(12):e0261002. doi 10.1371/journal.pone.0261002
3. Savola P. et al. Somatic mutations in clonally expanded cytotoxic T lymphocytes in patients with newly diagnosed rheumatoid arthritis. *Nat Commun*. 2017;8:15869. doi 10.1038/ncomms15869
4. Yang X., Garner L.I., Zvyagin I.V. et al. Autoimmunity-associated T cell receptors recognize HLA-B*27-bound peptides. *Nature*. 2022;612(7941):771-777. doi 10.1038/s41586-022-05501-7
5. Bolotin D.A., Poslavsky S., Mitrophanov I. et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. 2015;12(5):380-381. doi 10.1038/nmeth.3364
6. Mamedov I.Z., Britanova O.V., Zvyagin I.V. et al. Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Front Immunol*. 2013;4:456. doi 10.3389/fimmu.2013.00456
7. Koboldt D.C., Chen K., Wylie T. et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. 2009;25(17):2283-2285. doi 10.1093/bioinformatics/btp373
8. Shugay M., Bagaev D.V., Zvyagin I.V. et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res*. 2018;46(D1):D419-D427. doi 10.1093/nar/gkx760
9. Komech E.A., Koltakova A.D., Barinova A.A. et al. TCR repertoire profiling revealed antigen-driven CD8+ T cell clonal groups shared in synovial fluid of patients with spondyloarthritis. *Front Immunol*. 2022;13:973243. doi 10.3389/fimmu.2022.973243
10. Kanehisa M., Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi 10.1093/nar/28.1.27

doi 10.18699/bgrs2024-4.2-17

Low-frequency variants in a cohort of Russian primary immunodeficiency patients

Petrusenko Y.[1, 2]*, Chekanov N.[1], Musharova O.[1], Monakhova A.[1], Klimuk E.[1], Severinov K.[1, 2]

[1] *"Biotech campus" LLC, Moscow, Russia*

[2] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

** YPetrusenko@biotc.ru*

Key words: primary immunodeficiency; rare genetic variants; disease-causing mutations; whole genome sequencing; oligogenic model of inheritance

Motivation and Aim: Primary immunodeficiencies (PIDs) represent a heterogeneous group of inborn diseases varying by severity, manifestation age, clinical features, and genetic background [1]. In Russia, the mutational landscape of PID patients, especially of patients with sporadic immunodeficiencies, is understudied [2]. In this work, we conducted whole-genome sequencing (WGS) of a Russian PID cohort, classified genetic variants by clinical significance, and evaluated the percentage of novel and known pathogenic mutations. As a result, known genes enriched with rare mutations as well as several candidate genes were revealed.

Methods and Algorithms: DNA was extracted from blood samples provided by Saint-Petersburg Pasteur Institute Medical Centre and sequenced on an DNBSEQ-T7 MGI instrument. Sequence short reads were mapped to the human reference genome (hg38) using BWA. Variant calling was performed using DeepVariant [3] and SNP annotation with VEP (using CADD, Polyphen, SpliceAI etc plugins) [4]. The American College of Medical Genetics and Genomics (ACMG) guidelines were used to categorize variants [5]. The study group included 140 patients falling into several groups of diagnoses: Common variable immunodeficiency (CVID), combined immunodeficiency (CID), Bruton agammaglobulinemia (XLA), Selective IgA deficiency (sIgAD), Transient Hypogammaglobulinemia of Infancy (THI), Specific antibody deficiency, Lymphoproliferative syndrome, C1-INH-HAE, Unspecified immunodeficiency, etc. Informed consents were obtained for all the patients in accordance with the ethical standards.

Results: Germinal genetic variants in a panel of 620 genes associated with PID and other hereditary pathologies were selected in 140 patients. To reveal primarily monogenic cases of PID, we concentrated on truncating genetic changes (nonsense, frameshifts), missense mutations, and annotated splice sites in these genes. Variants with different levels of confidence were proposed based on clinical data. We concentrated on rare variants with population frequency in gnomAD and Russian population cohort of the "Biotech campus" LLC database of < 1 %. Most of detected rare variants were classified as variants of uncertain significance (VUS). 11 previously known pathogenic variants and 21 new likely pathogenic variants were identified (Table 1).

Table 1. Preliminary diagnoses and genes in which likely causal variants have been found

Preliminary diagnoses	Novel likely pathogenic variants	Known pathogenic/likely pathogenic variants
CVID	*NFKB1*, *SOCS1*, *PLCG2*	
Unspecified immunodeficiency	*SBDS*, *CD8A*	*SBDS*
CID	*TLR3*	*AIRE*, *CARD11*
C1-INH-HAE, type 1	*SERPING1*	*SERPING1*
XLA	*BTK*	*BTK*
Lymphoproliferative syndrome	*CTLA4*, *FAS*, *CARD11*	*GATA2*
Immune dysregulation	*RIPK1*	
Ataxia	*ATM*	*ATM*

Possible genetic causes of the disease were found in 34 patients. Most of them represented familial cases. In one patient with "unspecified immunodeficiency" (novel and known, presumably, in compound) mutations responsible for the Shwachman-Diamond syndrome were identified [6]. Another interesting case was a mutation in CD8A that was previously considered a VUS. In this patient, marked decrease in CD8+ T cells was observed, thereby indicating that this mutation could be pathogenic [7]. Analysis of genomic coding regions in 30 sporadic cases of CVID revealed disease-causing mutations in known genes in 4 (13 %) patients. The relatively low proportion of monogenic findings can suggest oligogenic model of inheritance in most cases [8]. We therefore considered the enrichment of rare variants in our genes panel. Several novel variants in *MAPK8* were detected, consistent with its role as a PID candidate gene [9].
Conclusion: This study highlights the potential of whole-genome sequencing to expand identification of novel variants and genes involved in PID. We evaluated frequencies of rare variants in a heterogenic cohort of Russian patients with immune system disorders, categorized them by genes, functionally predicted impacts, and clinical significance. Identified variants can be good candidates for future functional studies to establish their role in pathogenesis.
Funding: The study is supported by "Biotech campus" LLC.

References

1. Tangye S.G. et al. Human Inborn Errors of Immunity: 2022 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol.* 2022;42(7):1473. doi 10.1007/s10875-022-01289-3
2. Mukhina A.A. et al. Primary Immunodeficiencies in Russia: data from the National registry. *Front Immunol.* 2020;11:1491. doi 10.3389/fimmu.2020.01491
3. Poplin R. et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol.* 2018;36(10):983-987. doi 10.1038/nbt.4235
4. McLaren W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1):122. doi 10.1186/s13059-016-0974-4
5. Richards S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi 10.1038/gim.2015.30
6. Boocock G.R.B. et al. Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat Genet.* 2003;33(1):97-101. doi 10.1038/ng1062
7. Mancebo E. et al. Gly111Ser mutation in CD8A gene causing CD8 immunodeficiency is found in Spanish Gypsies. *Mol Immunol.* 2008;45(2):479-484. doi 10.1016/j.molimm.2007.05.022
8. Bisgin A., Sonmezler O., Boga I., Yilmaz M. The impact of rare and low-frequency genetic variants in common variable immunodeficiency (CVID). *Sci Rep.* 2021;11:8308. doi 10.1038/s41598-021-87898-1
9. de Valles-Ibáñez G. et al. Evaluating the genetics of common variable immunodeficiency: monogenetic model and beyond. Front Immunol. 2018;9:636. doi 10.3389/fimmu.2018.00636

doi 10.18699/bgrs2024-4.2-18

Generation of three iPSC lines from fibroblasts of a patient with Cohen syndrome

Pristyazhnyuk I.[1], Minina J.[1], Voinova V.[2, 3], Safonova M.[3], Lagarkova M.[4], Menzorov A.[1, 5]*

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Veltischev Research and Clinical Institute for Pediatrics and Pediatric Surgery of the Pirogov Russian National Research Medical University, Moscow, Russia*

[3] *The Mental Health Research Center, Moscow, Russia*

[4] *Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia*

[5] *Novosibirsk State University, Novosibirsk, Russia*

* *menzorov@bionet.nsc.ru*

Key words: iPSCs; Cohen syndrome

Motivation and Aim: Cohen syndrome is a rare autosomal recessive genetic disorder characterized by developmental delay, intellectual disability, microcephaly, neutropenia, dimorphisms, obesity, and ophthalmological and autistic spectrum disorders [1]. It is caused by homozygous or compound heterozygous variants in *VPA13B* gene. More than 200 pathological variants of the *VPS13B* gene in more than 1000 patients are currently registered [2]. The protein is localized on the Golgi complex. It has many domains and the function of some of them is still not clear. VPS13B is involved in lipid transport between cell membranes and is crucial for vesicular trafficking and autophagy. Previously we generated Cohen syndrome patient-specific induced pluripotent stem cells (iPSCs). The iPSC-derived neural stem cells and neurons had numerous ultrastructural abnormalities [3]. In this study, we produced iPSCs from fibroblasts of another Cohen patient that had *VPS13B* gene compound heterozygous variants with unknown clinical significance. A study of unknown variants may clarify the functions of VPS13B as well as the functional roles of its domains.

Methods and Algorithms: Generation and cultivation of the iPSC lines and methods of analysis were described previously [4]. Cell culture was performed at the Collective Center of ICG SB RAS "Collection of Pluripotent Human and Mammalian Cell Cultures for Biological and Biomedical Research" (https://ckp.icgen.ru/cells/; http://www.biores.cytogen.ru/brc_cells/collections/ICG_SB_RAS_CELL) supported by the state project FWNR-2022-0019.

Results: We have generated three iPSC lines from fibroblasts of the Cohen patient with previously unreported compound heterozygous variants (8:g.99766811A>G and 8:g.99859429G>A, HG38). Episomal vectors (Epi5) were used to reprogram fibroblasts. Sanger sequencing confirmed the presence of variants. The cell lines iCS-MCF3-1, iCS-MCF3-3, and iCS-MCF3-5 had normal karyotypes and expressed pluripotency markers OCT4, NANOG, and SSEA-4. Pluripotency was shown by analysis of embryoid bodies; they expressed gene markers of all three germ layers.

Conclusion: These iPSC lines could be used for the study of VPS13B functions and the impact of its variants on the cell function.

Funding: The study was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2021-1063).

References

1. Seifert W., Holder-Espinasse M., Kühnisch J. et al. Expanded mutational spectrum in Cohen syndrome, tissue expression, and transcript variants of *COH1*. *Hum Mutat.* 2009;30(2):E404-E420. doi 10.1002/humu.20886
2. Douzgou S., Petersen M.B. Clinical variability of genetic isolates of Cohen syndrome. *Clin Genet.* 2011;79(6):501-506. doi 10.1111/j.1399-0004.2011.01669.x
3. Shnaider T.A., Khabarova A.A., Morozova K.N. et al. Ultrastructural abnormalities in induced pluripotent stem cell-derived neural stem cells and neurons of two Cohen syndrome patients. *Cells.* 2023;12(23):2702. doi 10.3390/cells12232702
4. Pristyazhnyuk I.E., Meshcheryakov N.I., Nikitina T.V. et al. Generation of induced pluripotent stem cell line iTAF15Xsk4 from fibroblasts of a patient with microdeletion at Xq24. *Russ J Dev Biol*. 2023;54:358-364. doi 10.1134/S1062360423060073

doi 10.18699/bgrs2024-4.2-19

Spatial transcriptional landscape of early-onset tongue cancer

Prostakishina E.[1, 2]*, Patysheva M.[2], Kolegova E.[2], Menyailo M.[2], Khozyainova A.[2], Fedorova I.[2], Arora R.[3], Choinzonov E.[2], Kulbakin D.[2], Bose P.[3], Denisov E.[2]

[1] *Tomsk State University, Tomsk, Russia*

[2] *Cancer Research Institute, Tomsk National Research Medical Center, RAS, Tomsk, Russia*

[3] *Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada*

* *elprostakishina@yandex.ru*

Key words: tongue cancer; young adults; immune microenvironment; spatial transcriptomics

Motivation and Aim: Tongue squamous cell cancer (TSCC) is characterized by aggressive local invasion and metastasis [1]. The incidence of TSCC among individuals under 45 years of age is growing every year [2]. The etiological factors and pathogenetic mechanisms of TSCC in young adults are poorly understood due to the lack of association with tobacco, alcohol consumption, and HPV [3]. Here, we aimed to investigate the TSCC landscape in patients younger than 45 years of age using spatial transcriptomics.

Methods and Algorithms: The study included 9 TSCC patients with $T_{2\text{-}3}N_{0\text{-}1}M_0$ divided into two groups: under 45 years (n=5) and older than 45 years (n=4). FFPE tumor samples from 4 HPV-negative patients under 45 years and fresh frozen tumor samples from 5 HPV-negative patients (1 patient aged < 45 years and 4 patients aged > 45 years) were profiled using the 10x Genomics Visium Spatial Gene Expression platform. Sequencing was performed on a GenoLab M instrument (GeneMind Biosciences, China) and a NovaSeq 6000 (Illumina, USA). Data integration was performed using the Harmony program package. Seraut package and Enrichr tool were used for clusters identification and gene expression characterization.

Results: A total of 9 clusters were identified with a log2FC > 0.50 and adjusted p-value < 0.0001, which were divided into eight major cell types: two clusters of tumor cells, lymphoid and myeloid immune cells, cancer-associated fibroblasts, muscle, normal epithelium, and salivary gland cells. In early-onset TSCC, tumor clusters were enriched by increased expression of mitogen activated kinase (MAPK) pathway genes: *EGR1*, *IL1B*, *MAP2K1*, *MAP2K3*, *FOS*, *FOSB*, *JUN*, *JUNB*, *FOSL1*, and *FOSL2* (log2FC > 0.57 and adjusted p-value < 0.0001) compared to late-onset cancer. At the same time, there was a decreased expression of oxidative phosphorylation genes (*ATF5PF*, *SDHD*, *GBP5*, *NDUFA3*, *UQCR11*, *COX7B*, *S100A7*, log2FC > 0.48 and adjusted p-value < 0.0001) and increased expression of glycolysis marker *PFKFB3* (log2FC > 0.74 and adjusted p-value < 0.0001) in young adults TSCC. The border between tumor and normal tissue in the TSCC of young adults was enriched with vascular mimicry with increased expression of JAK-STAT signaling pathway genes (*TGA5*, *ITGB1*, *MAP2K3*, *STAT3*, *MYC*, *EPHA2*, *FN1*, *MMP14*, *COL1A1*; log2FC > 0.61 and adjusted p-value < 0.0001). In addition, regions of vascular mimicry were enriched with tumor-associated macrophage markers: *CD68*, *SERPINB2* and *OSM*.

Conclusion: TSCC in young adults has distinct spatial transcriptomic features compared to late-onset tongue cancer including enrichment of MAPK and JAK-STAT signaling pathways and TAMs and vascular mimicry at the invasive front as well as downregulation of oxidative phosphorylation and upregulation of glycolysis.
Funding: The study is supported by the Russian Science Foundation (No. 22-15-00308).

References

1. Bagan J., Sarrion G., Jimenez Y. Oral cancer: Clinical features. *Oral Oncol*. 2010;46(6):414-417. doi 10.1016/j.oraloncology.2010.03.009
2. Dos Santos Costa S.F., Brennan P.A., Gomez R.S. et al. Molecular basis of oral squamous cell carcinoma in young patients: Is it any different from older patients? *J Oral Pathol Med.* 2018;47(6):541-546. doi 10.1111/jop.12642
3. da Silva Souto A.C., Vieira Heimlich F., Lima de Oliveira L., Bergmann A., Dias F.L., Spíndola Antunes H., de Melo A.C., Thuler L.C.S., Cohen Goldemberg D. Epidemiology of tongue squamous cell carcinoma: A retrospective cohort study. *Oral Dis.* 2023;29(2):402-410. doi 10.1111/odi.13897

doi 10.18699/bgrs2024-4.2-20

Inframe indels in human proteins: abundance and effect prediction

Ramensky V.[1, 2, 3]*

[1] *Institute for Advanced Research on Artificial Intelligence and Intelligent Systems, Lomonosov Moscow State University, Moscow, Russia*

[2] *National Medical Research Center for Therapy and Preventive Medicine of the Ministry of Healthcare of Russian Federation, Moscow, Russia*

[3] *Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

* *ramensky@gmail.com*

Key words: inframe indel; clinical significance; effect prediction; gnomAD; ClinVar

Motivation and Aim: Short insertions and deletions that do not induce frameshifts in protein (inframe indels) are of considerable interest because their functional and clinical effects remain largely unexplored and range from severe disease to no phenotypic manifestations.

Methods and Algorithms: Currently, databases contain more than 300 thousand human indels with known clinical significance and/or population frequency. Since large-scale experimental assessment of indel clinical and/or functional effects is not feasible, computational approaches that provide effect prediction are very relevant. Currently, both "classical" bioinformatics methods and novel deep learning approaches are used for effect prediction.

Results: Methods based on large protein language models (e. g., ESM1b) and not using databases of indels with known clinical relevance for training seem to be the most interesting [1], since they are free from some of the disadvantages of classical supervised learning-based methods (e. g., SIFT, CADD, Provean) [2]. We developed a method for predicting the effect of indels, based on the annotation of the sequence of the protein under study and deep alignments with homologs. We independently evaluated the performance of this approach and a number of other predictive methods [3] for indels of known clinical significance or population frequency and compared the results.

Conclusion: From the practical point of view, the most promising approach is a combination of several independent prediction methods

References

1. Brandes N., Goldman G., Wang C.H. et al. Genome-wide prediction of disease variant effects with a deep protein language model. *Nat Genet.* 2023;55:1512-1522. doi 10.1038/s41588-023-01465-0
2. Cannon S., Williams M., Gunning A.C., Wright C.F. Evaluation of in silico pathogenicity prediction tools for the classification of small in-frame indels. *BMC Med Genomics.* 2023;16:36. doi 10.1186/s12920-023-01454-6
3. Fan X., Pan H., Tian A. et al. SHINE: protein language model-based pathogenicity prediction for short inframe insertion and deletion variants. *Brief Bioinform.* 2023;24:bbac584. doi 10.1093/bib/bbac584

doi 10.18699/bgrs2024-4.2-21

FUT2 gene polymorphism in samples of Tuvinians and Russians of Eastern Siberia

Tabikhanova L.E.[1]*, Osipova L.P.[1], Churkina T.V.[1], Lichman D.V.[1], Voronina E.N.[2], Filipenko M.L.[2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

* *tabikhan@bionet.nsc.ru*

Key words: Tuvinians; Russians of Eastern Siberia; genetic polymorphism; real-time PCR; *FUT2* (*G772A*, *rs602662*)

Motivation and Aim: The *FUT2* gene encodes α-1,2-fucosyl-transferase, an enzyme that modifies oligosaccharides in the intestine, which are a substrate for microflora. The state of the *G772A rs602662* polymorphic locus affects FUT2 enzyme activity: the *-772G* variant is secretory, while the *-772A* allele is non-secretory. FUT2 secretory status differentially affects susceptibility to a number of infections: the FUT2 "secretor" phenotype (allele *-772G*) is associated with susceptibility to viral infections, including influenza, Epstein-Barr virus, SARS-COV 2, and *Helicobacter pylori* [1]. "Nonsecretory" FUT2 phenotype (allele *-772A*) is associated with susceptibility to bacterial pathogens, including *Streptococcus pneumoniae*, *Streptococcus pneumoniae*, *Neisseria meningitidis*, *Haemophilus infl uenzae*, and *Salmonella typhimurium*. [2]. These differences are reflected in susceptibility to multifactorial diseases. For example, chronic pancreatitis is associated with a nonsecretory phenotype [3]. Approximately 20 % of the global population carries the non-functional *-772A* allele, and there are interpopulation differences. In East Asia, less than 2 % of the population has *-772A* in their genotype, while in European countries the proportion is 30-50 % [4]. In the literature, we have not encountered data on the frequencies of polymorphic variants of the *FUT2* gene in indigenous populations of Siberia, so research in this area remains relevant.

Methods and Algorithms: Two Siberian samples served as the material for the present research. Persons of Tuvinian nationality having no ancestors with foreign ethnic backgrounds and living in the city of Kyzyl of Tuva Republic were included in the Tuvinian group (N = 308). Russians from Eastern Siberia, whose ancestors lived in the settlements of Zabaykalsky Krai and the Irkutsk Region for several generations, were included in the second sample (N=153).

The frequencies of polymorphic variants due to the single nucleotide substitution *G772A (rs602662)* of the *FUT2* gene were determined by the real-time polymerase chain reaction method. Population allele frequencies of polymorphic variants were determined based on the observed genotype frequencies. The match between the empirically observed genotype frequency distribution and the expected theoretical distribution in Hardy-Weinberg equilibrium was examined using $\chi 2$ Pearson test (the equilibrium holds at $p>0.05$). Significance of differences in allele frequencies between the studied samples was determined using $\chi 2$ test with Yates's correction for continuity; the results were considered statistically significant at $p<0.05$. The obtained parameters were compared with those in the literature.

Results: Genotype distribution for polymorphic locus of the *FUT2 G772A (rs602662)* in samples of Tuvinians and Russians from Eastern Siberia is presented in Table 1.

Table 1. Genotype distribution for *FUT2 G772A (rs602662)* in samples of Tuvinians and Russians from Eastern Siberia

Population			Tuvinians	Russians from Eastern Siberia
–772G/A, rs602662	Genotype distribution *n*	*G/G*	237	66
		G/A	61	70
		A/A	10	17
	N, *ppl*		308	153
	p (H–W)		0.581	0.808

Note. *N* is the sample size, *p (H-W)* is the probability of deviation from the Hardy–Weinberg equilibrium.

Genotype distribution matched the Hardy-Weinberg equilibrium for both polymorphic loci. The frequency of the non-secretory allele *-772A* in the sample of Russians from Eastern Siberia (34 %) corresponds to the data for Caucasoids. The frequency of allele in the sample of Tuvinians were statistically significantly lower than in samples of Russians from Eastern Siberia and Caucasoid groups described in the literature [4]. On the other hand, frequencies of the *-772A* variant in Tuvinian sample significantly increased if compared to a number of East Asian groups. In accordance with the general geographical gradient of the distribution of this variant, its frequency in the Tuvinian sample is in an intermediate position between Caucasoid and East Asian populations. The same trend was revealed by us earlier in studies of polymorphisms of other functionally important genes [5].

Conclusion: Thus, the allele frequencies of the polymorphic locus *FUT2 G772A (rs602662)* in Tuvinians and Russians of Eastern Siberia were determined for the first time. It was shown that in the Russian sample the frequency of the non-secretory allele *-772A* is within the frequency range for other Caucasoid populations. The frequency of *-772A* in the Tuvinian sample is in an intermediate position between Caucasoids and East Asian populations. The statistically significantly reduced frequency of the non-secretory allele *-772A* of the *FUT2* gene in the Tuvinian population may indicate an increased risk of gastrointestinal pathologies in case of *H. pylori* infection, compared to the Russians. At the same time, the Tuvinian population probably has a reduced risk of chronic pancreatitis. However, to test this hypothesis, additional medical genetic studies are needed in various populations with a large sample size, as well as the study of the frequencies of allelic variants, not only of *FUT2*, but also of other functionally significant genes.

Funding: The study was carried out within the state task of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (No. FWNR-2022-0021).

References

1. Currier R., Payne D., Staat M., Selvarangan R., Shirley S. et al. Innate susceptibility to norovirus infections infl uenced by FUT2 genotype in a United States pediatric population. *Clin Infect Dis*. 2015;60:1631-1638. doi10.1093/cid/civ165
2. Goto Y., Obata T., Kunisawa J., Sato S., Ivanov I. et al. Innate lymphoid cells regulateintestinal epithelial cell glycosylation. *Science*. 2014;345:125400. doi 10.1126/science.1254009
3. Weiss F., Schurmann C., Guenther A., Ernst F., Teumer A. et al. Fucosyltransferase 2 (FUT2) non-secretor status and blood group B are associated with elevated serum lipase activity in asymptomatic subjects, and an increased risk for chronic pancreatitis: a genetic association study. *Gut*. 2015;64(4):646-656. doi 10.1136/gutjnl-2014–306930
4. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65. doi 10.1038/nature11632
5. Tabikhanova L.E., Osipova L.P., Churkina T.V., Kovalev S.S., Filipenko M.L., Voronina E.N. Increased Frequencies of the –174G and –572C IL6 Alleles in Populations of Indigenous Peoples of Siberia Compared to Russians. *Mol Biol*. 2023;57(2):350-359. doi 10.1134/S002689332302019X

doi 10.18699/bgrs2024-4.2-22

Aberrations of the placental methylome and embryonic death

Vasilyev S.A.*, Demeneva V.V., Tolmacheva E.N., Vasilyeva O.Yu., Shevtsov D.G., Lushnikov I.V., Fonova E.A., Sazhenova E.A., Nikitina T.V., Lebedev I.N.

Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia

** stanislav.vasilyev@medgenetics.ru*

Key words: DNA methylation; methylome; placenta; embryonic death; spontaneous abortions

Motivation and Aim: DNA methylation is one of the key mechanisms responsible for the implementation of the individual development program. The purpose of this study is to identify the potential role of multiple abnormalities in the methylome of extraembryonic tissues in the death of human embryos at early stages of development.

Methods and Algorithms: The methylation level was assessed using reduced representation bisulfite sequencing (RRBS) in chorionic villi of 8 spontaneous abortions of the first trimester of pregnancy with 45,X karyotype and 7 induced abortions with normal 46,XX and 46,XY karyotypes. Using STR analysis, the parental origin of the X chromosome was established in spontaneous abortions 45,X. Also, we analyzed 8 spontaneous abortions of the first trimester of pregnancy with normal karyotype with high and low level of LINE-1 methylation using RRBS.

Moreover, we analyzed the methylation profile of the LINE-1 retrotransposon using targeted bisulfite massive parallel sequencing in pairs of spontaneous abortions from the same families (n = 21 pairs).

Results: When spontaneous abortions with monosomy X of both maternal and paternal origin were compared to induced abortions with a normal karyotype of 46,XX and 46,XY, a number of differentially methylated CpG sites were found. Lamina-associated domains and subtelomeric regions of the genome were shown to have an enrichment of differentially methylation sites. A comparison with induced abortions with the 46,XX and 46,XY karyotypes revealed 831 and 254 differentially methylated genes (DMG), respectively. Genes of proteins with intrinsically disordered domains were enriched among genes with differentially methylation promoter regions in all spontaneous abortions with monosomy X. The two comparison groups shared 48 DMGs, of which 21 genes are important in normal placental and embryonic development and whose dysregulation is linked to preeclampsia and embryonic death. Furthermore, among the common DMGs with differences in the methylation index of ≥40 %, there are genes linked to the function of the Golgi apparatus and calcium channels in the cell membrane, which can cause pathologies in the excretory and cardiovascular systems in live births with Turner syndrome.

In spontaneous abortions with a normal karyotype, RRBS revealed large-scale anomalies in gene methylation, among which developmental protein genes were enriched.

Previously, we found an increased level of LINE-1 methylation in spontaneous abortions with an aneuploid karyotype. In this study, the level of methylation of the LINE-1 retrotransposon was found to be significantly correlated in spontaneous abortions from the same families ($n = 21$ pairs, $R = 0.71$, $p = 0.0003$). For pairs of spontaneous abortions with a normal karyotype ($n = 10$ pairs), the correlation remains significant ($R = 0.74$, $p = 0.015$), while for pairs of spontaneous abortions in which one of the abortions has

aneuploidy, the correlation becomes statistically insignificant (n = 8 pairs, R = 0.64, p = 0.08). If we exclude spontaneous abortions of half-siblings (n = 5 pairs), then the correlation also remains significant (n = 16 pairs, R = 0.67, p = 0.005). At the same time, no significant differences were found between the methylation level in all first abortions and all subsequent abortions in pairs (47.9 ± 6.9 % versus 43.9 ± 6.8 %, p = 0.062).
The role of the LINE-1 retrotransposon in the placenta is unclear. It was shown that among the genes capable of using LINE-1 promoters, genes expressed in the brain and placenta are enriched. In addition, retrotransposons from the L1PA2 subfamily may provide their antisense promoters as alternative promoters of long noncoding RNAs in the placenta. It is possible that LINE-1 elements incapable of retrotransposition gradually begin to be used by the genome as separate functional elements (primarily promoters) for nearby genes. In this case, an increase in the level of LINE-1 methylation can lead to disruption of the expression of nearby genes and disrupt the development of the placenta.
Conclusion: Aberrant methylation of genes involved in placentation, proliferation, and cell differentiation has been found in spontaneous miscarriages with monosomy X and normal karyotype. These disorders are likely to be the cause of the embryo lethality. The detected correlation between the levels of methylation of the LINE-1 retrotransposon in the chorionic villi of spontaneous abortions from the same families may be a consequence of common maternal factors affecting embryos and associated with their death, or indirect evidence of a transgenerational effect on the level of LINE-1 methylation of at least one from parents.
Funding: The study is supported by Russian science foundation (No. 23-15-00341).

4 Симпозиум «Эволюционная, популяционная и медицинская геномика/генетика человека: компьютерные и экспериментальные подходы»

Symposium “Evolutionary, population and medical genomics/genetics of human: computational and experimental approaches”

doi 10.18699/bgrs2024-4.3-01

Оценка эффективности работы инструментов для функциональной аннотации результатов GWAS с использованием симулированных данных

Барбитов Ю.*, Чангалиди А., Глотов А.
ФГБНУ «НИИ акушерства, гинекологии и репродуктологии им. Д.О. Отта», Санкт-Петербург, Россия
** barbitoff@bk.ru*

Ключевые слова: GWAS; сложные признаки; генетический вариант; симулированные данные

Мотивация и цель: Полногеномный анализ ассоциаций (genome-wide association studies, GWAS) является важнейшим подходом для поиска генетических основ сложных признаков человека и других организмов. За последние десятилетия были проведены сотни успешных GWAS-исследований; в то же время аннотация и интерпретация обнаруженных ассоциаций до сих пор остаются сложной задачей. Разработка эффективных методов для функциональной аннотации результатов GWAS затрудняется отсутствием готовых данных с известным набором причинных генетических вариантов и биологических процессов, вовлеченных в развитие изучаемого признака. Целью настоящей работы являлось создание метода для симуляции GWAS-данных с предварительно заданным набором причинных вариантов и/или процессов, а также применение этого инструмента для оценки точности работы инструментов для поиска обогащения наборов генов в результатах GWAS.
Методы и алгоритмы: В ходе работы был разработан новый инструмент, bioGWAS, позволяющий проводить симуляции реалистичных наборов GWAS-данных. bioGWAS был реализован в виде биоинформатического протокола на языке Snakemake, включающего в себя стадии симуляции генотипов (с использованием HAPGEN2 [1]), фенотипов (при помощи PhenotypeSimulator [2]) и результатов анализа ассоциаций между ними. Разработанный инструмент был применен для оценки эффективности работы часто используемых методов для поиска обогащенных наборов генов в результатах GWAS: MAGMA [3] и Pascal [4], а также разработанного ранее в нашей научной группе инструмента LSEA. Исходный код и текущая версия bioGWAS доступны по адресу https://github.com/TohaRhymes/bioGWAS/ [5].
Результаты: Разработанный инструмент для симуляции GWAS-данных продемонстрировал свою высокую эффективность для генерации результатов GWAS с заданным набором причинных вариантов и наборов генов, в которых расположены данные варианты (показатель F1 при обнаружении заданных вариантов в качестве значимых ассоциаций выше 0.95). Использование данных, сгенерированных с помощью bioGWAS, для оценки эффективности идентификации заданных наборов генов инструментами MAGMA, Pascal и LSEA показал значительное влияние размера целевого набора генов на чувствительность обнаружения обогащения, а также позволил сделать вывод о значительном преимуществе метода LSEA над аналогами (чувствительность LSEA составила от

83 до 100 % для разных наборов генов, чувствительность других инструментов варьировала в диапазоне от 5 до 100 %).
Выводы: Разработанный в ходе работы инструмент bioGWAS расширяет арсенал подходов для создания и оценки эффективности методов аннотации результатов GWAS, а инструмент LSEA позволяет с высокой чувствительностью обнаруживать биологические процессы, вовлеченные в формирование сложных признаков.
Финансирование: Исследование поддержано Министерством науки и высшего образования Российской Федерации (проект «Многоцентровая научно-исследовательская коллекция биоресурсов «Репродуктивное здоровье человека»» (контракт № 075-15-2021-1058 от 28 сентября 2021 г.), а также стипендией по системной биологии, полученной Барбитовым Ю.А. в 2021–2023 гг.

Assessing the performance of tools for functional annotation of GWAS results using simulated data

Barbitoff Y.*, Changalidis A., Glotov A.
Research Institute of Obstetrics, Gynecology and Reproductology named after D.O. Ott, St. Petersburg, Russia
** barbitoff@bk.ru*

Key words: GWAS; complex traits; genetic variant; simulated data

Motivation and Aim: Genome-wide association study (GWAS) is among the most important approaches for finding the genetic basis of complex traits in humans and other organisms. Hundreds of successful GWAS studies have been conducted over the past decades; however, annotation and interpretation of the detected associations is still a challenging task. The development of efficient methods for functional annotation of GWAS results is complicated by the lack of readily available datasets with a known set of causal genetic variants and biological processes involved in the development of the studied trait. The goal of this work was to create a method for simulating GWAS datasets with a predefined set of causal variants and/or processes, and to apply this tool to evaluate the accuracy of tools for finding gene set enrichment in GWAS results.
Methods and Algorithms: In this work, we developed a new tool, bioGWAS, to enable simulations of realistic GWAS datasets. bioGWAS was implemented as a bioinformatic pipeline using the Snakemake toolkit; it includes simulation of genotypes (using HAPGEN2 [1]), phenotypes (using PhenotypeSimulator [2]), and summary statistics of their association. The developed tool was applied to evaluate the performance of commonly used methods for finding enriched gene sets in GWAS results: MAGMA [3] and Pascal [4], as well as the LSEA tool developed earlier in our research group. The source code and the current version of bioGWAS are available at https://github.com/TohaRhymes/bioGWAS/ [5].
Results: Our tool demonstrated its high efficiency for generating GWAS results with a given set of causal variants and gene sets in which these variants are located (F1 score for detecting predefined variants as significant associations is more than 0.95). We next used the data generated by bioGWAS to evaluate the efficiency of identification of given gene sets by MAGMA, Pascal, and LSEA tools. We found a significant effect of the size of the target gene set on the sensitivity of enrichment analysis; at the same time, results

showed that the LSEA method has a significant advantage over its competitors (LSEA sensitivity ranged from 83 to 100 % for different gene sets, while the sensitivity of other tools ranged from 5 to 100 %).

Conclusion: The bioGWAS tool expands the arsenal of approaches for development and evaluation of the performance of GWAS results annotation methods, while the LSEA tool allows for highly sensitive detection of biological processes involved in complex trait formation.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (project "Multicenter research bioresource collection" "Human Reproductive Health" contract No. 075-15-2021-1058 from September 28, 2021), as well as by Systems Biology Fellowship to Y.A.B. in 2021–2023.

Список литературы/References

1. Su Z., Marchini J., Donnelly P. HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics*. 2011;27(16):2304-2305. doi 10.1093/bioinformatics/btr341
2. Meyer H.V., Biner E. PhenotypeSimulator: A comprehensive framework for simulating multi-trait, multi-locus genotype to phenotype relationships. *Bioinformatics*. 2018;34(17):2951-2956. doi 10.1093/bioinformatics/bty197
3. de Leeuw C.A., Mooij J.M., Heskes T., Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219. doi 10.1371/journal.pcbi.1004219
4. Lamparter D. et al. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput Biol*. 2016;12(1):e1004714. doi 10.1371/journal.pcbi.1004714
5. Changalidis A.I. et al. bioGWAS: A Simple and Flexible Tool for Simulating GWAS Datasets. *Biology (Basel)*. 2023;13(1):10. doi 10.3390/biology13010010

doi 10.18699/bgrs2024-4.3-02

Шкалы генетического риска для комплексных фенотипов и их применение к российской популяции

Зайченока М.[1]*, Жарикова А.А.[2, 3], Вяткин Ю.В.[2, 4], Киселева А.В.[2], Ершова А.И.[2], Сотникова Е.А.[2], Мешков А.Н.[2, 5, 6, 7], Раменский В.Е.[2, 3, 8], Драпкина О.М.[2]

[1] *Московский физико-технический институт (национальный исследовательский университет), Долгопрудный, Московская область, Россия*

[2] *Национальный медицинский исследовательский центр терапии и профилактической медицины Минздрава России, Москва, Россия*

[3] *Московский государственный университет им. М.В. Ломоносова, факультет биоинженерии и биоинформатики, Москва, Россия*

[4] *Новосибирский государственный университет, Новосибирск, Россия*

[5] *Национальный медицинский исследовательский центр кардиологии им. академика Е.И. Чазова Минздрава России, Москва, Россия*

[6] *Медико-генетический научный центр им. академика Н.П. Бочкова, Москва, Россия*

[7] *Российский национальный исследовательский медицинский университет им. Н.И. Пирогова Минздрава России, Москва, Россия*

[8] *Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем МГУ им. М.В. Ломоносова, Москва, Россия*

** marija.zaicenoka@gmail.com*

Ключевые слова: шкалы генетического риска; российская популяция; липидный профиль крови; семейная гиперхолестеринемия; атеросклероз

Мотивация и цель: Сердечно-сосудистые заболевания являются лидирующей причиной смертности в России [1]. Шкалы генетического риска (ШГР) позволяют оценивать индивидуальную предрасположенность развития комплексного (мультифакторного) заболевания до его клинических проявлений. Липидный профиль крови, включающий в себя уровнень липопротеинов низкой (ЛПНП) и высокой плотности (ЛПВП), уровень триглицеридов (ТГ) и уровень общего холестерина (ОХС), тесно связан с развитием различных сердечно-сосудистых заболеваний – ишемической болезни сердца, атеросклероза и геморрагического инсульта. В последние годы было проведено множество разных полногеномных исследований, посвященных поиску ассоциаций вариантов генома с уровнями липидов крови [2, 3], на основе которых можно строить различные ШГР. Целью настоящей работы является оценка применимости опубликованных ранее ШГР для фенотипов липидного профиля к российской популяции, поиск оптимального метода применения данных ШГР и оценка применимости ШГР в клинической практике.

Методы и алгоритмы: В исследовании используются три выборки:

- популяционная выборка из Ивановской области, состоящая из 1673 человек (624 мужчины, 1049 женщин), собранная в ходе исследования ЭССЕ-РФ [4]. Медианный возраст выборки составляет 50 лет. Выборка была секвенирована с использованием таргетной панели, содержащей 242 гена и более 2000 участков генома с клинически важными однонуклеотидными вариантами, входящими в опубликованные ШГР;

- популяционная выборка из Ивановской области, состоящая из 817 человек (346 мужчин, 417 женщин), собранная во время исследования ЭССЕ-РФ [4]. Медианный возраст выборки составляет 47 лет. Выборка была секвенирована с использованием таргетной панели, содержащей 217 генов и более 18 000 участков генома с клинически важными однонуклеотидными вариантами;
- выборка больных семейной гиперхолестеринемией (СГХС), состоящая из 372 человек (107 мужчин, 265 женщин). Медианный возраст выборки составляет 57 лет. Выборка была секвенирована с помощью полногеномного секвенирования.

Для каждого участника были измерены показатели липидного профиля – ЛПНП, ЛПВП, ОХС и ТГ. В рамках субисследования АТЕРОГЕН-Иваново участникам исследования ЭССЕ-РФ по Ивановской области было проведено УЗИ сонных и бедренных артерий. Данные обследования имеются для 61.5 % участников выборки из Ивановской области. Имеются данные о стенозе, высоте атеросклеротических бляшек, их количество, толщина интима-медиа.

Для построения ШГР были использованы данные из трех различных статей [2, 3, 6], включающие в себя от 1308 до 2 432 587 вариантов для различных липидных фенотипов.

Рассматривались три способа применения ШГР:

- использование ШГР «как есть» на основе исходных оценок эффекта всех значимых вариантов;
- поиск оптимальной ШГР с помощью подхода C+T (clumping & thresholding), реализованного в методе PRSice-2 [7];
- расчет ШГР с использованием LDpred-2 [8].

Используя имеющиеся выборки, был проведен расчет R^2 (процент объясняемой дисперсии) полученных ШГР для фенотипов ЛПНП, ЛПВП, ТГ, ОХС.

Полученные шкалы были использованы для классификации больных/здоровых СГХС, а также для изучения связи наличия патогенного моногенного варианта и значения ШГР. Была произведена оценка возможности использования ШГР для липидного профиля для предсказания результатов исследования УЗИ бедренных и сонных артерий.

Результаты: В рамках работы было проанализировано по 20 ШГР для четырех показателей липидного профиля для Ивановской и Вологодской выборок соответственно: 12 шкал «как есть» и 8 разработанных методами PRSice-2 и LDpred-2. Результаты R^2 варьировали от 3.5 до 12 % (что совпадает с литературными данными) для различных фенотипов, наилучшие результаты были показаны при построении ШГР методом LDpred-2.

Показано, что распределения ШГР для ЛПНП и ОХС у больных СГХС и здорового контроля из Ивановской области значимо различаются. Более того, распределения значений ШГР для ОХС и ЛПНП значимо различаются между больными – носителями патогенных вариантов в генах *LDLR*, *APOB* и *PCSK9*, и больными неносителями.

Выводы: В рамках работы была оценена эффективность применения ШГР, построенных на основе данных для популяций, отличных от популяций Российской Федерации. Также была оценена возможность клинической применимости ШГР для различных заболеваний, в частности, СГХС и атеросклероза.

Polygenic risk scores for complex phenotypes and their application to the Russian population

Zaicenoka M.[1]*, Zharikova A.A.[2, 3], Vyatkin Yu.V.[2, 4], Kiseleva A.V.[2], Ershova A.I.[2], Sotnikova E.A.[2], Meshkov A.N.[2, 5, 6, 7], Ramensky V.E.[2, 3, 8], Drapkina O.M.[2]

[1] *Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Moscow region, Russia*

[2] *National Medical Research Center for Therapy and Preventive Medicine of the Ministry of Health of Russia, Moscow, Russia*

[3] *Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow, Russia*

[4] *Novosibirsk State University, Novosibirsk, Russia*

[5] *National Medical Research Center of Cardiology named after academician E.I. Chazov of the Ministry of Health of Russia, Moscow, Russia*

[6] *Medical Genetic Research Center named after academician N.P. Bochkov, Moscow, Russia*

[7] *Russian National Research Medical University named after N.I. Pirogov of Ministry of Health of Russia, Moscow, Russia*

[8] *Institute for Advanced Research on Artificial Intelligence and Intelligent Systems, Lomonosov Moscow State University, Moscow, Russia*

* *marija.zaicenoka@gmail.com*

Key words: genetic risk scales; Russian population; blood lipid profile; familial hypercholesterolemia; atherosclerosis

Motivation and Aim: Cardiovascular diseases are the leading cause of mortality in Russia [1]. Polygenic risk scores (PRS) allow one to assess an individual's susceptibility to the development of a complex (multifactorial) disease before its clinical manifestations. The blood lipid profile, including low-density lipoprotein (LDL) and high-density lipoprotein (HDL) levels, triglyceride (TG) levels and total cholesterol (TC) levels, is closely associated with the development of various cardiovascular diseases – coronary heart disease, atherosclerosis and hemorrhagic stroke. In recent years, many different genome-wide studies have been conducted to search for associations of genome variants with blood lipid levels [2, 3], on the basis of which various PRS can be constructed. The purpose of this work is to assess the applicability of previously published PRS for lipid profile phenotypes to the Russian population, to find the optimal method for using PRS data and to assess the applicability of PRs in clinical practice.

Methods and Algorithms: This study uses three samples:

- A population sample from the Ivanovo region, consisting of 1,673 people (624 men, 1,049 women), collected during the ESSE-RF study [4]. The median age of the sample is 50 years. The sample was sequenced using a targeted panel containing 242 genes and more than 2,000 genomic regions with clinically important single nucleotide variants included in the published PRS;
- A population sample from the Ivanovo region, consisting of 817 people (346 men, 417 women), collected during the ESSE-RF study [4]. The median age of the sample is 47 years. The sample was sequenced using a targeted panel containing 217 genes and more than 18,000 genomic regions with clinically important single nucleotide variants;

- A sample of patients with familial hypercholesterolemia (FH), consisting of 372 people (107 men, 265 women). The median age of the sample is 57 years. The sample was sequenced using whole genome sequencing.

For each participant, lipid profile indicators were measured – LDL, HDL, total cholesterol and TG. As part of the ATEROGEN-Ivanovo substudy, participants in the ESSE-RF study in the Ivanovo region underwent ultrasound of the carotid and femoral arteries. Survey data are available for 61.5 % of sample participants from the Ivanovo region. There is data on stenosis, the height of atherosclerotic plaques, their number, and intima-media thickness.

To construct the PRS, data from three different articles were used [2, 3, 6], including from 1,308 to 2,432,587 variants for various lipid phenotypes.

Three methods of using PRS were considered:

- Application of the PRS "as is" based on initial estimates of the effect of all significant variants;
- Finding the optimal PRS using the C+T (clumping & thresholding) approach implemented in the PRSice-2 method [7];
- Calculation of PRS using LDpred-2 [8].

Using the available samples, R^2 was calculated (percentage of explained variance) of the obtained PRS for the LDL, HDL, TG, and TC phenotypes.

The resulting scales were used to classify patients/healthy patients with FH, as well as to study the relationship between the presence of a pathogenic monogenic variant and the value of PRS. The feasibility of using lipid profile PRS to predict femoral and carotid artery ultrasound findings was assessed.

Results: As part of the work, 20 PRS were analyzed for 4 lipid profile indicators for the Ivanovo and Vologda samples, respectively: 12 "as is" scales and 8 developed by the PRSice-2 and LDpred-2 methods. R^2 of resulting PRS varied from 3.5 to 12 % (which coincides with the literature data) for various phenotypes, the best results were shown when constructing the PRS using the LDpred-2 method.

It was shown that the distributions of PRS for LDL and TC in patients with FH and healthy controls from the Ivanovo region are significantly different. Moreover, the distributions of PRS values for total cholesterol and LDL differ significantly between patients who are carriers of pathogenic variants in the genes *LDLR*, *APOB* and *PCSK9*, and sick non-carriers.

Conclusion: As part of the work, the effectiveness of using the PRS, built on the basis of data for populations other than the populations of the Russian Federation, was assessed. The possibility of clinical applicability of PRS for various diseases, in particular FH and atherosclerosis, was also assessed.

Список литературы/References

1. Боровкова Н.Ю., Токарева А.С., Савицкая Н.Н., Крисанова К.И., Курашин В.К., Одинцов Г.А. Современное состояние проблемы сердечно-сосудистых заболеваний в Нижегородском регионе: возможные пути снижения смертности. *Российский кардиологический журнал*. 2022;27(5):5024. doi 10.15829/1560-4071-2022-5024
2. Selvaraj M.S., Li X., Li Z. et al. Whole genome sequence analysis of blood lipid levels in >66,000 individuals. *Nat Commun*. 2022;13:5995. doi10.1038/s41467-022-33510-7
3. Xu Y., Ritchie S.C., Liang Y. et al. An atlas of genetic scores to predict multi-omic traits. *Nature*. 2023;616:123-131. doi 10.1038/s41586-023-05844-9
4. Бойцов С.А., Драпкина О.М., Шляхто Е.В., Конради А.О., Баланова Ю.А., Жернакова Ю.В., Метельская В.А., Ощепкова Е.В., Ротарь О.П., Шальнова С.А. Исследование ЭССЕ-РФ (Эпидемиология сердечно-сосудистых заболеваний и их факторов риска в регионах Российской

Федерации). Десять лет спустя. *Кардиоваскулярная терапия и профилактика*. 2021;20(5):3007. doi 10.15829/1728-8800-2021-3007
5. Ершова А.И., Балахонова Т.В., Мешков А.Н., Куценко В.А., Яровая Е.Б., Шальнова С.А., Лищенко Н.Е., Новикова А.С., Александрова Е.Л., Шутемова Е.А., Белова О.А., Рачкова С.А., Бойцов С.А., Драпкина О.М. Распространенность атеросклероза сонных и бедренных артерий среди населения Ивановской области: исследование АТЕРОГЕН-Иваново. *Кардиоваскулярная терапия и профилактика*. 2021;20(5):2994. doi 10.15829/1728-8800-2021-2994
6. Willer C.J., Schmidt E.M., Sengupta S. et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274-1283. doi 10.1038/ng.2797
7. Choi S.W., O'Reilly P.F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*. 2019;8(7):giz082. doi 10.1093/gigascience/giz082
8. Privé F., Arbel J., Vilhjálmsson B.J. LDpred2: better, faster, stronger. *Bioinformatics*. 2021;36(22-23):5424-5431. doi 10.1093/bioinformatics/btaa1029

doi 10.18699/bgrs2024-4.3-03

Являются ли импутированные генотипы альтернативой данных экзомного секвенирования при анализе редких вариантов?

Зоркольцева И.В.*, Аксенович Т.И., Цепилов Я.А.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
* *zor@bionet.nsc.ru*

Ключевые слова: анализ ассоциаций на основе генов; биобанк Великобритании; бинарные признаки; количественные признаки; белок-кодирующие варианты; некодирующие варианты

Мотивация и цель: Полногеномные исследования ассоциаций (GWAS) – эффективный способ изучения ассоциаций генотип-фенотип. Этот подход позволил получить значительное количество надежных ассоциаций для многих сложных признаков и распространенных заболеваний человека. Но результаты GWAS объясняют лишь малую долю наследуемости. Для объяснения феномена «потерянной наследуемости» было предложено несколько подходов, в том числе анализ редких вариантов [1]. Такой анализ стал возможен с развитием технологий секвенирования экзомов. Однако широкое использование генотипов, полученных с помощью секвенирования, в настоящее время ограничено относительно небольшим количеством секвенированных образцов в биобанках и отсутствием свободного доступа к персональным генотипам. Напротив, результаты GWAS для тысяч признаков и заболеваний, полученные на сотнях тысяч образцов, сегодня находятся в свободном доступе (например, каталог GWAS) и могут быть использованы для поиска ассоциаций на основе генов (gene-based analysis), который является основным инструментом для анализа редких (частота минорного аллеля, MAF < 0.01) и ультраредких (MAF < 0.001) генетических вариантов [2, 3]. Целью данного исследования является проверка гипотезы о том, что импутированные генотипы могут являться дополнительным источником информации для анализа редких вариантов [4]. Для этого мы сравнили результаты анализа редких вариантов на основе генов с использованием импутированных и секвенированных генотипов, полученных на выборке одних и тех же участников проекта UK Biobank.

Методы и алгоритмы: Исследование проведено на выборке европейцев из проекта UK Biobank (N = 188 236). Использованы секвенированные (11334327 аутосомных SNVs, MAF < 0.01) и импутированные генотипы (4258875 аутосомных SNPs, INFO > 0.8, missingness rate < 0.02 и 5×10^{-5} < MAF < 0.01).

Были проанализировали 20 бинарных (случай-контроль) и 27 количественных признаков. Одноточечный анализ ассоциаций для каждого признака проводили с помощью пакета fastGWA-GLMM, версия 1.94.0 beta [5].

Для анализа ассоциаций на основе генов использовали суммарные статистики (z-scores и размеры эффекта betas) и матрицы корреляций между генотипами всех вариантов в пределах гена. Анализ проводили для каждого из трех наборов вариантов: 1) несинонимичные (LoF + missense), 2) все кодирующие, 3) все

некодирующие варианты в гене. Перед анализом секвенированных данных генотипы ультраредких вариантов (число минорных аллелей, MAC ≤ 10) были объединены в один вариант [6]. Использовались два регрессионных метода, основанных на суммарных статистиках: SKAT-O [7] и PCA [8], реализованных в пакете sumFREGAT R [3]. Результаты методов SKAT-O и PCA были объединены с помощью агрегированного омнибус-теста Коши, ACAT-O [9]. Анализ ограничивался белок-кодирующими генами с не менее чем двумя вариантами, имеющими суммарные статистики.
Результаты: Результаты анализа представлены в табл. 1.

Таблица 1. Сравнение числа сигналов ассоциации (p-*value* < 2.5×10^{-6}) между импутированными (Имп) и секвенированными (Сек) генотипами для трех наборов вариантов в гене

Тип признака	Несинонимичные			Кодирующие			Некодирующие		
	Всего	Имп	Сек	Всего	Имп	Сек	Всего	Имп	Сек
Бинарный	25	2	24	20	3	18	18	14	4
Количественный	579	257	538	648	334	594	1610	1530	318

Количество сигналов ассоциации, выявленных для бинарных признаков, было меньше, чем для количественных, вследствие низкой распространенности бинарных признаков (min = 0.061, mean = 0.097, max = 0.292). Большее число ассоциаций было выявлено для несинонимичных и кодирующих вариантов при использовании секвенированных данных по сравнению с импутированными. Для количественных признаков секвенированные данные были примерно в два раза эффективнее импутированных данных по числу идентифицированных генов. Для некодирующих вариантов импутированные генотипы позволили выявить больше сигналов ассоциации, чем секвенированные данные. Для количественных признаков это соотношение составило 4.8. Некоторые сигналы ассоциации были обнаружены с помощью импутированных генотипов, но не обнаружены с помощью секвенированных.
Выводы: Мы показали, что импутированные генотипы могут быть использованы для анализа редких кодирующих вариантов, особенно когда секвенированные данные недоступны. Для некодирующих областей генов импутированные данные даже более информативны, чем данные, полученные с помощью секвенирования экзома. Мы считаем, что импутированные генотипы являются эффективным и при этом доступным источником данных для анализа редких вариантов.
Финансирование: Исследование поддержано грантом РНФ № 23-25-00209.

Are imputed genotypes an alternative to exome sequencing data when analysing rare variants?

Zorkoltseva I.V.*, Axenovich T.I., Tsepilov Y.A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
* *zor@bionet.nsc.ru*

Key words: gene-based association analysis; UK biobank; binary traits; quantitative traits; protein-coding variants; non-protein-coding variants

Motivation and Aim: Genome-wide association studies (GWAS) are a powerful tool for studying genotype-phenotype associations. This methodology has generated a myriad of robust associations for a range of complex traits and common human diseases. However, GWAS results explain only a small fraction of heritability. Several approaches have been proposed to explain the phenomenon of missing heritability, including the analysis of rare variants [1]. Such analysis has became possible with the development of exome sequencing technologies. However, the widespread use of sequencing-derived genotypes is now limited by the relatively small number of sequenced samples in biobanks and the lack of free access to personal genotypes. On the contrary, GWAS results for thousands of traits and diseases from hundreds of thousands of samples are now freely available (e. g. GWAS catalogue) and can be used for gene-based association analysis, which is the main tool for analysing rare (minor allele frequency, MAF < 0.01) and ultra-rare (MAF < 0.001) genetic variants [2, 3]. The aim of this study is to test the hypothesis that imputed genotypes can be an additional source for rare variant analysis [4]. To do this, we compare the results of gene-based association analysis of rare variants using imputed and sequenced genotypes from the same participants in the UK Biobank project.
Methods and Algorithms: The study was conducted using individuals of white European ancestry from the UK Biobank with both exome sequenced (11334327 autosomal SNVs with MAF < 0.01) and imputed genotypes (4258875 autosomal SNPs with INFO > 0.8, missingness rate < 0.02 and 5×10^{-5} < MAF < 0.01) (N = 188236).
We analysed 20 binary (case-control) and 27 quantitative traits. A single variant association analysis for each trait was performed using the fastGWA-GLMM tool, version 1.94.0 beta [5].
For gene-based association analysis, summary statistics (z scores and effect sizes) and matrices of correlations between genotypes of all variants within a gene were used.
A gene-based association analysis was performed for each of three different variant annotations: LoF + missense, all coding and all non-coding variants in the gene. Before testing the sequenced data for each variant annotation, the genotypes of ultra-rare variants (minor allele count, MAC ≤ 10) were collapsed into a single variant [6]. Two regression-based methods based on summary statistics were used: SKAT-O [7] and PCA [8], implemented in the sumFREGAT R package [3]. The results of SKAT-O and PCA were combined using the aggregated Cauchy omnibus test, ACAT-O [9]. Analysis was restricted to protein-coding genes with at least two variants that had the summary statistics.
The Bonferroni adjusted significance level for the total number of genes (20000) was defined as 2×10^{-6}.
Results: The results of the analyses are presented in Table 1.

Table 1. Comparison of the number of signals (p-value < 2.5×10^{-6}) between imputed (imput) and sequenced (seq) genotypes for binary and quantitative traits. A comparison is shown for each annotation analysed

Trait	Lof+missence			Coding			Non-coding		
	All	Imput	Seq	All	Imput	Seq	All	Imput	Seq
Binary	25	2	24	20	3	18	18	14	4
Quantitave	579	257	538	648	334	594	1610	1530	318

The number of association signals observed for binary traits was lower than for quantitative traits due to the low prevalence of binary traits (min = 0.061, mean = 0.097,

max = 0.292). More associations were identified for non-synonymous and coding variants when using sequenced data compared to imputed data. For quantitative traits, sequenced data were about twice as efficient as imputed data in terms of gene number. For non-coding variants, the imputed genotypes detected more association signals than the sequenced data. For quantitative traits, this ratio was 4.8. There are some association signals that were detected with the imputed genotypes but not with the sequenced ones. *Conclusion*: We showed that the imputed genotypes can be useful for analysing coding variants, especially when sequenced data are not available. For non-coding regions of genes, imputed data are even more informative than data obtained by exome sequencing. We believe that imputed genotypes are a reasonable source for analysing rare variants.
Funding: The study is supported by the Russian Science Foundation (RSF) grant No. 23-25-00209.

Список литературы/References

1. Wainschtein P., Jain D., Zheng Z. et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat Genet*. 2022;54(3):263-273. doi 10.1038/s41588-021-00997-7
2. de Leeuw C.A., Mooij J.M., Heskes T., Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219. doi 10.1371/journal.pcbi.1004219
3. Svishcheva G.R., Belonogova N.M., Zorkoltseva I.V., Kirichenko A.V., Axenovich T.I. Gene-based association tests using GWAS summary statistics. *Bioinformatics*. 2019;35(19):3701-3708. doi 10.1093/bioinformatics/btz172
4. Wuttke M., Konig E., Katsara M.A., Kirsten H., Farahani S.K., Teumer A. et al. Imputation-powered whole-exome analysis identifies genes associated with kidney function and disease in the UK Biobank. *Nat Commun*. 2023;14(1):1287. doi 10.1038/s41467-023-36864-8
5. Jiang L., Zheng Z., Fang H., Yang J. A generalized linear mixed model association tool for biobank-scale data. *Nat Genet*. 2021;53(11):1616-1621. doi 10.1038/s41588-021-00954-4
6. Zhou W., Bi W., Zhao Z., Dey K.K., Jagadeesh K.A., Karczewski K.J. et.al. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat Genet*. 2022;54:1466-1469. doi 10.1038/s41588-022-01178-w
7. Li B., Leal S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet.* 2008;83(3):311-321. doi 10.1016/j.ajhg.2008.06.024
8. Wang K., Abbott D. A principal components regression approach to multilocus genetic association studies. *Genet Epidemiol*. 2008;32(2):108-118. doi 10.1002/gepi.20266
9. Liu Y., Chen S., Li Z., Morrison A.C., Boerwinkle E., Lin X. ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am J Hum Genet.* 2019;104(3):410-421. doi 10.1016/j.ajhg.2019.01.002

doi 10.18699/bgrs2024-4.3-04

Новая стратегия условного анализа генных ассоциаций

Свищёва Г.Р.*, Белоногова Н.М., Цепилов Я.А., Аксенович Т.И.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
** gulsvi@bionet.nsc.ru*

Ключевые слова: модель со случайными эффектами; анализ ассоциаций; условное распределение

Мотивация и цель: Анализ ассоциаций на основе генов (gene-based association analysis, GBAA) тестирует совместный эффект на признак всех однонуклеотидных полиморфизмов (SNP) внутри гена. Как правило, материалом для этого анализа являются z-статистики, полученные в результате полногеномного анализа ассоциаций (GWAS), и матрицы корреляций между генотипами SNP внутри генов. Как для GWAS, так и для GBAA обязательным элементом является условный анализ, позволяющий исключать влияние внегенных SNP, находящихся в неравновесии по сцеплению с SNP внутри гена. Одним из известных инструментов условного анализа является метод GCTA-COJO [1]. Этот метод корректирует z-статистики для каждого SNP внутри гена с учетом внегенных сигналов. Скорректированные z-статистики вместе с матрицами корреляций затем используются в GBAA. Простой и быстрой альтернативой COJO для GBAA является метод полигенного прунинга (polygenic pruning, PP) [2]. В его основе лежит исключение из рассмотрения внутригенных SNP, высоко коррелирующих с более значимо ассоциированными SNP вне гена. Оба этих метода изменяют число или значения z-статистик, участвующих в GBAA, не меняя матрицы генетических корреляций. Недостатком COJO являются вычислительные проблемы при обращении матрицы, а недостатком PP – потенциальное снижение мощности из-за удаления некоторых SNP. Здесь мы представляем новый метод условного анализа TauCOR, лишенный этих ограничений. Стратегия TauCOR состоит в том, что он не меняет значений z-статистик, а корректирует матрицы корреляций между генотипами SNP внутри гена. Такой подход позволяет избежать многих вычислительных проблем.

Методы и алгоритмы: Рассмотрим выборку генотипированных индивидов с измеренным фенотипом, для которой проведен GWAS. Для каждого гена мы определяем набор SNP внутри гена (g) и набор SNP из региона вокруг гена (r). Для этих наборов известны векторы z-статистик, z_r и z_g, и их корреляционные матрицы, U_g и U_r, соответственно. Кроме того, известна матрица корреляций между внутри- и внегенными SNP, U_{gr}. Задача состоит в том, чтобы скорректировать матрицу корреляций U_g, необходимую для GBAA, таким образом, чтобы учесть z_r, U_r и U_{gr}. В новом методе мы предполагаем, что эффекты SNP в регионе вокруг гена носят независимый случайный характер и распределены нормально с общим вкладом в изменчивость признака τ, а для эффектов SNP гена характер (случайный/фиксированный) на данном этапе не важен. Сначала мы анализируем SNP в регионе вокруг гена, чтобы оценить их вклад. Для этого мы строим распределение z_r с нулевым вектором матожиданий и корреляционной матрицей $U_r + \tau U_r U_r$. С помощью метода максимума правдоподобия, построенного для этого распределения, оцениваем τ. Затем анализируем ген интереса, рассмотрев для него

два потенциальных источника ассоциативного сигнала: SNP самого гена и SNP соседнего региона с уже известным вкладом τ. В этом случае нулевой гипотезой является предположение об отсутствии сигналов в гене, вызванных непосредственно SNP гена. В соответствии с нулевой гипотезой мы корректируем корреляционную матрицу U_g так, чтобы она отражала сигналы внегенных SNP, в результате чего получаем $U_g + \tau U_{gr} U_{gr}^T$. Далее выполняем GBAA с помощью любого из тестов, используя в качестве входных данных первичные результаты GWAS и скорректированные матрицы корреляций. Наиболее популярные GBAA-тесты представлены в [3].

Результаты: TauCOR был протестирован на наборе генов ‘золотого стандарта’, т. е. генов, эффект которых на признак однозначно установлен. Было отобрано 28 генов ‘золотого стандарта’ на 13 признаках. Мы использовали z-статистики из UK Biobank для этих признаков и сравнили новый метод с методами COJO и PP.

Для каждого из 28 генов ‘золотого стандарта’ были сформированы локусы в пределах ±1 Мб от границ гена. Всего в этих локусах содержалось 394 гена помимо генов из списка ‘золотого стандарта’. В противовес истинно-положительным генам ‘золотого стандарта’ в данном исследовании мы условно рассматриваем остальные гены в локусах как истинно-отрицательные для данных признаков. Число таких генов на локус варьировало от 1 до 59 со средним 14.6.

GBAA проводили с помощью R-пакета sumFREGAT [4]. Для каждого гена SNP были отфильтрованы по частоте минорного аллеля $\geq 10^{-4}$, аннотированы с использованием VEP и организованы в три набора: некодирующие, кодирующие и несинонимические SNP. Для каждого набора SNP был выполнен тест ACAT-O, объединяющий результаты тестов SKATO и PCA. Всего, с учетом перекрывания анализируемых локусов, было протестировано 84 набора SNP для истинно-положительных генов и 1211 наборов для истинно-отрицательных генов. При проведении условного анализа для каждого гена рассматривали регион вокруг гена ±5 Мб для COJO и PP и ±1.5 Мб для TauCOR.

Прежде всего, мы провели GBAA и отобрали значимо ассоциированные наборы SNP для каждого гена, показавшие P-значения $\leq 2.5 \times 10^{-6}$. Среди 84 тестов, выполненных на наборах SNP для генов ‘золотого стандарта’, было 29 значимых, а среди 1211 тестов, выполненных на наборах SNP для генов из их окружения, было 97 значимых. Для этих значимых тестов проводили условный анализ. Фиксировали число тестов, прошедших условный анализ, для генов ‘золотого стандарта’, чтобы оценить чувствительность метода, и число тестов для генов из соседних регионов, чтобы оценить специфичность методов. Результаты приведены в табл. 1.

Таблица 1. Показатели эффективности различных методов

	GBAA	COJO	PP	TauCOR
Гены золотого стандарта	29	12	15	16
Соседние гены	97	17	22	5
Чувствительность		0.41	0.52	0.55
Специфичность		0.82	0.77	0.95

Как видно, число значимых тестов, полученных новым методом для генов ‘золотого стандарта’, лишь немного больше, чем у существующих методов. Это свидетельствует о его чувствительности, сравнимой с чувствительностью других

методов условного анализа. Однако специфичность нового метода оказалась значительно выше, чем у существующих. TauCOR показал ложно-положительный результат только в пяти случаях, тогда как другие методы показывали такой результат в четыре раза чаще.
Выводы: Стратегия TauCOR продемонстрировала сравнимую чувствительность условного анализа и значительно более высокую его специфичность по сравнению с COJO и PP. Метод может быть успешно использован для повышения точности анализа ассоциаций на генном уровне.
Финансирование: Исследование выполнено при финансовой поддержке гранта Российского научного фонда (РНФ) № 23-25-00209.

A new strategy for conditional gene-based association analysis

Svishcheva G.R.*, Belonogova N.M., Tsepilov Y.A., Axenovich T.I.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** gulsvi@bionet.nsc.ru*

Key words: random effects model; association analysis; conditional distribution

Motivation and Aim: Gene-based association analysis (GBAA) tests the joint effect of all single nucleotide polymorphisms (SNPs) within a gene on a trait. As a rule, the input data for this analysis are z-statistics obtained by genome-wide association analysis (GWAS) and correlation matrices between genotypes of SNPs within genes. For both GWAS and GBAA, conditional analysis is a mandatory step to exclude the influence of extra-gene SNPs that are in linkage disequilibrium with SNPs within a gene. One of the well-known conditional analysis tools is the GCTA-COJO method [1]. This method adjusts the z-statistics for each SNP within a gene to account for the extra-gene signals. These adjusted z-statistics together with correlation matrices are then used in GBAA. A simple and fast alternative to COJO for GBAA is the polygenic pruning (PP) method [2]. PP is based on the exclusion of intra-gene SNPs that are highly correlated with more significantly associated SNPs outside the gene. Both of these methods change the number or values of z-statistics involved in GBAA without changing the matrix of genetic correlations. The disadvantage of COJO is the computational problems in inverting the matrix, and the disadvantage of PP is the potential reduction in power due to the removal of some SNPs. Here we present a new conditional analysis method, TauCOR, devoid of these limitations. The strategy of TauCOR is that it does not change the values of the z-statistics, but adjusts the correlation matrices between genotypes of SNPs within a gene. This approach avoids many computational problems.
Methods and Algorithms: Consider a sample of genotyped individuals with a measured phenotype for which a GWAS was performed. For each gene, we define a set of SNPs within the gene (g) and a set of SNPs from the region around the gene (r). For these sets, we know the vectors of z-statistics, z_g and z_r, and their correlation matrices, U_g and U_r, respectively. Furthermore, the correlation matrix between intra- and extra-gene SNPs, U_{gr}, is known. The task is to adjust the correlation matrix, U_g, required for GBAA so as to account for z_r, U_r and U_{gr}.
In the new method, we assume that the effects of SNPs in the region around the gene are independently random and normally distributed with an common contribution to trait variability τ, and for the effects of SNPs in the gene, the type (random/fixed) is not

important at this stage. We first analyse the SNPs in the region around the gene to estimate their contribution. To do this, we construct a z_r distribution with zero mean vector and correlation matrix $U_r + \tau U_r U_r$. Using the maximum likelihood method constructed for this distribution, we estimate τ. We then analyse the gene of interest by considering two potential sources of association signal for it: SNPs of the gene itself and SNPs of a neighbouring region with an already known contribution τ. In this case, the null hypothesis is the assumption that there are no signals in the gene caused directly by intragenic SNPs. Under this null hypothesis, we adjust the correlation matrix U_g so that it reflects signals from extra-gene SNPs, resulting in $U_g + \tau U_{gr} U_{gr}^T$. We then perform GBAA with either test, using the primary GWAS results and the adjusted correlation matrices as input data. The most popular GBAA tests are presented in [3].

Results: TauCOR was tested on a set of 'gold standard' genes, i.e. genes whose influence on a particular trait has been unambiguously established. Twenty-eight 'gold standard' genes were selected for 13 traits. We used z-statistics from UK Biobank for these traits and compared the new method with COJO and PP methods.

For each of the 28 'gold standard' genes, loci were generated within ±1 Mb of the gene boundaries. These loci contained a total of 394 genes in addition to the 'gold standard' genes. As a contrast to the true-positive genes from 'gold standard', in this study we conditionally consider the remaining genes in the loci as true-negatives for these traits. The number of such genes per locus ranged from 1 to 59, with an average of 14.6.

GBAA was performed using the R package sumFREGAT [4]. For each gene, SNPs were filtered by minor allele frequency $\geq 10^{-4}$, annotated using VEP and divided into three sets: non-coding, coding and non-synonymous SNPs. For each set of SNPs, the ACAT-O test was performed combining the results of the SKATO and PCA tests. A total of 84 SNP sets for true-positive genes and 1211 SNP sets for true-negative genes were tested, taking into account the overlap of the loci analysed. In the conditional analysis for each gene, we considered a region around the gene in size of ±5 Mb for COJO and PP and ±1.5 Mb for TauCOR.

First of all, we performed GBAA and selected significantly associated SNP sets for each gene that showed P-values $\leq 2.5\times10^{-6}$. Among the 84 tests performed on SNP sets for 'gold standard' genes, there were 29 significant tests, and among the 1,211 tests performed on SNP sets for genes from their surroundings, there were 97 significant tests. Conditional analysis was performed for these significant tests. To assess the sensitivity of the methods, the number of tests that passed the conditional assay for 'gold standard' genes was counted, and to assess the specificity of the methods, the number of tests for genes from neighbouring regions was counted. The results are summarized in Table 1.

Table 1. Effectiveness indicators of various methods

	GBAA	COJO	PP	TauCOR
'Gold standard' genes	29	12	15	16
Neighbouring genes	97	17	22	5
Sensitivity		0.41	0.52	0.55
Specificity		0.82	0.77	0.95

As can be seen, the number of significant tests obtained by the new method for 'gold standard' genes is only slightly higher than that of the existing methods. This indicates its sensitivity comparable to that of other conditional analysis methods. However, the specificity of the new method was significantly higher than that of existing methods.

TauCOR showed false-positive results in only 5 cases, while other methods showed such results four times more often.

Conclusion: The TauCOR's strategy demonstrated comparable sensitivity and significantly higher specificity of conditional analysis in comparison to COJO and PP. The method can be successfully used to improve the accuracy of association gene-based analyses.

Funding: The study was financially supported by the Russian Science Foundation (RSF) grant No. 23-25-00209.

Список литературы/References

1. Yang J., Lee S.H., Goddard M.E., Visscher P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76-82
2. Belonogova N.M., Zorkoltseva I.V., Tsepilov Y.A., Axenovich T.I. Gene-based association analysis identifies 190 genes affecting neuroticism. *Sci Rep.* 2021;11(1):2484
3. Svishcheva G.R., Belonogova N.M., Zorkoltseva I.V., Kirichenko A.V., Axenovich T.I. Gene-based association tests using GWAS summary statistics. *Bioinformatics.* 2019;35(19):3701-3708
4. Belonogova N.M., Svishcheva G.R., Kirichenko A.V., Zorkoltseva I.V., Tsepilov Y.A., Axenovich T.I. sumSTAAR: a flexible framework for gene-based association studies using GWAS summary statistics. *PLOS Comput Biol.* 2022;18(6):e1010172

doi 10.18699/bgrs2024-4.3-05

Efficiency of a new multi-trait approach applied to back pain-related phenotypes

Belonogova N.M.[1]*, Elgaeva E.E.[1, 2], Zorkoltseva I.V.[1], Kirichenko A.V.[1], Svishcheva G.R.[1, 3], Freidin M.B.[4], Williams F.M.K.[5], Suri P.[6, 7, 8, 9], Axenovich T.I.[1], Tsepilov Y.A.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

[3] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

[4] *Department of Biology, School of Biological and Behavioral Sciences, Queen Mary University of London, London, UK*

[5] *Department of Twin Research and Genetic Epidemiology, King's College London, London, UK*

[6] *Seattle Epidemiologic Research and Information Center, VA Puget Sound Health Care System, Seattle, USA*

[7] *Division of Rehabilitation Care Services, Seattle, USA*

[8] *Clinical Learning, Evidence, and Research Center, University of Washington, Seattle, USA*

[9] *Department of Rehabilitation Medicine, University of Washington, Seattle, USA*

* *belon@bionet.nsc.ru*

Key words: chronic back pain; dorsalgia; intervertebral disc disorders; shared heritability; GWAS

Motivation and Aim: Back pain (BP) is a major contributor to disability worldwide. We conducted a cross-sectional study analyzing three BP-related phenotypes: chronic BP (CBP), dorsalgia and intervertebral disc disorders (IDD), with heritability estimated at 40–60 %. Less than half of the heritability is explained by common genetic variants identified by GWAS. More powerful methods of statistical analysis may offer additional insights. In this study, we utilized a novel multi-trait approach to merge three BP-related phenotypes: CBP, dorsalgia, and IDD, which have been previously used in BP studies. Dorsalgia is a phenotype based on electronic health records (EHR) and is defined by clinical diagnostic codes for back and neck pain, reflecting mainly the former. IDD is another phenotype related to BP and defined by EHR-based codes for intervertebral disc degeneration. These codes are typically used when BP is present, but not necessarily. There are high genetic correlations between BP-related phenotypes, estimated to be in the range of 87 to 92 % [1, 2], allowing for an approach that maximizes multi-trait heritability. The aim of this approach is to reduce the genetic heterogeneity of the traits and to identify a common genetic background for a number of traits that are genetically correlated [3].

Methods and Algorithms: The SHAHER framework aims to create a multi-trait phenotype from a set of genetically correlated traits and identify the genetic variants that control this trait [3]. The method assumes that the genetic background of each of the genetically correlated traits can be decomposed into two components: one common to all traits (the shared genetic impact or SGI), and the other specific to each trait. A new multi-trait phenotype, SGIT (the shared genetic impact trait), was created by linearly combining the original traits. The coefficients in this combination were defined by maximizing the heritability of SGIT. A GWAS of SGIT was performed using an

algorithm based on the summary statistics calculated for the original traits. The SHAHER method was previously described in our studies [2, 3].

Using imputed genotypes from the UK Biobank, we conducted a multi-trait gene-based association analysis of three BP-related phenotypes: CBP, dorsalgia and IDD. The first step of the analysis was performed using a discovery sample from the UK Biobank. Initially, the GWAS summary statistics for the three BP-related phenotypes were obtained. They were used for the calculation of the multi-trait (SGIT) summary statistics. Then we performed a gene-based association analysis (GBA) of the individual and multi-trait phenotypes. For all identified genes, conditional analysis (COJO) was applied. The second step of the analysis used a replication sample from the FinnGen database. We conducted a replication study by applying GBA to the 32 genes identified in the first step and using GWAS summary statistics for the two BP-related traits available on the FinnGen sample.

Results: We identified and replicated 16 genes associated with BP-related traits (Fig. 1). Seven of the detected genes, namely, *MIPOL1*, *PTPRC*, *RHOA*, *MAML3*, *JADE2*, *MLLT10* and *RERG*, have not been previously reported. The rest nine genes have been previously detected as associated with traits genetically correlated with BP or included in pathways associated with BP.

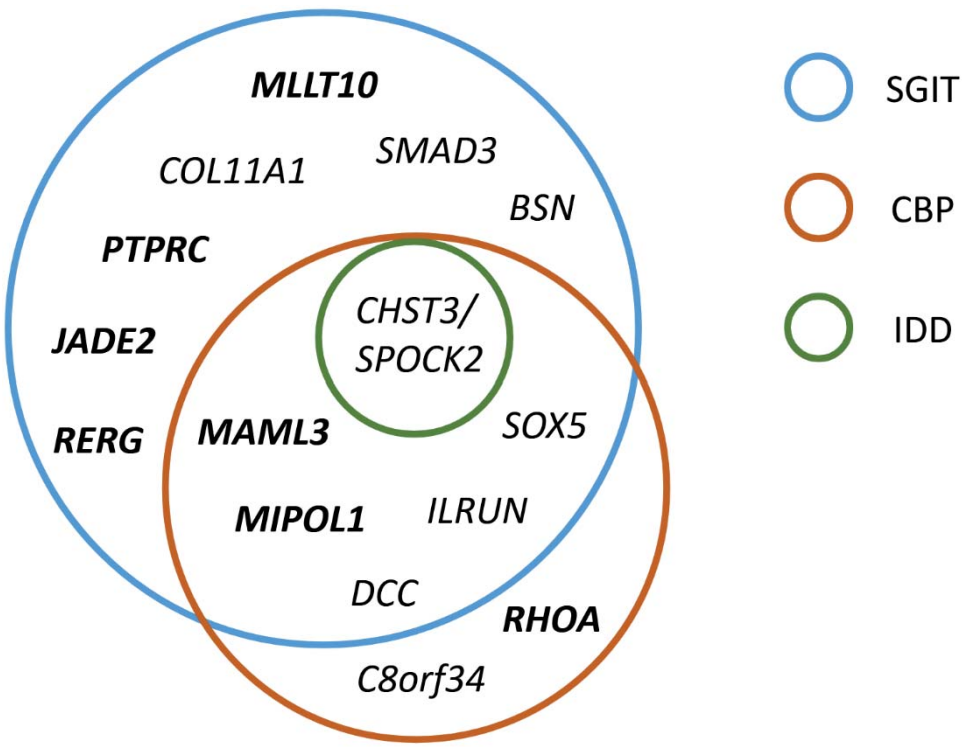


Fig. 1. Genes replicated on FinnGen data. Names of previously unreported genes are shown in bold

Conclusion: We conducted a genome-wide gene-based association analysis of BP-related phenotypes using imputed genotypes from a UK Biobank project and applying the SHAHER technique of multi-trait analysis [3].

SHAHER constructs a multi-trait phenotype as a linear combination of original traits, where the coefficients are estimated to maximize the heritability of the multi-trait phenotype. As a result, we identified seven genes associated with SGIT, which did not show a significant association with the original traits. Additional six genes were associated with both SGIT and CBP. These results confirm the effectiveness of the chosen strategy.

In total, we identified and replicated 16 genes significantly associated with BP-related traits. Thirteen genes were detected on the multi-trait phenotype that is in accordance with high genetic correlation between BP-related traits. Some genes have been previously described as associated with BP or with the genetically correlated traits, or as included in pathways associated with BP. Our results verify the role of these genes in BP-related traits and provide new insights into the genetics of back pain.

Funding: The work was supported by the Russian Science Foundation (RSF) grant No. 22-15-20037 and the Government of the Novosibirsk region.

Acknowledgments: This research has been conducted using the UK Biobank Resource under Applications No. 18219 and No. 59345. We want to acknowledge the participants and investigators of the FinnGen study.

References

1. Bjornsdottir G., Stefansdottir L., Thorleifsson G. et al. Rare *SLC13A1* variants associate with intervertebral disc disorder highlighting role of sulfate in disc pathology. *Nat Commun.* 2022;13(1):634. doi 10.1038/s41467-022-28167-1
2. Zorkoltseva I.V., Elgaeva E.E., Belonogova N.M., Kirichenko A.V. et al. Multi-trait exome-wide association study of back pain-related phenotypes. *Genes (Basel)*. 2023;14(10):1962. doi 10.3390/genes14101962
3. Svishcheva G.R., Tiys E.S., Elgaeva E.E. et al. A novel framework for analysis of the shared genetic background of correlated traits. *Genes (Basel).* 2022;13:1694. doi 10.3390/genes13101694

doi 10.18699/bgrs2024-4.3-06

Identification of the optimal reference panel for the imputation of genotype data from the Russian sample on the example of the RuDDS cohort

Berdnikova A.A.[1, 2]*, Tsepilov Y.A.[2], Nostaeva A.V.[1], Leonova O.N.[2, 3], Elgaeva E.E.[1, 2]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Priorov Central Institute of Traumatology and Orthopedic, Moscow, Russia*

* *a.berdnikova1@g.nsu.ru*

Key words: imputation; GWAS; SNP; genotyping; DNA microarray

Motivation and Aim: Imputation is a method that allows you to restore missing information about genetic variants that were not genotyped directly from DNA microarrays. Its main idea is to find haplotypes that are similar between the sample of interest (which is usually genotyped with DNA microarrays) and a reference sample (which has whole-genome sequencing data). Imputation of genotype data is an important step in genome wide association studies (GWAS) because it leads to a significant increase in the number of analyzed variants, which therefore improves the resolution of GWAS and enhances the comparability of the data obtained from diverse cohorts and/or different microarrays. In foreign studies, TOPMed reference panel has been considered as a gold standard for imputation, but the transferability of these recommendations to Russian samples is poorly studied. In this regard, the search for an optimal reference panel for imputation of Russian population data is an essential problem. In this study we aimed to test three most popular reference panels adapted for individuals of European descent (HRC, 1000 Genomes, TOPMed) to find an optimal reference panel for imputation of genotype data from Russian population.

Methods and Algorithms: We used genotype data, obtained utilizing DNA microarray with whole-genome coverage (Illumina Infinium Global Screening Array-24 v3.0 Kit) in a sample of self-defined Russian individuals from the Russian Disc Degeneration Study [1]. These data passed preliminary quality control resulting in 132 samples and 449604 SNPs. We then performed imputation of processed data using three reference panels: HRC (Version r1.1 2016) [2], 1000 Genomes Phase 3 (Version 5) [3], and TOPMed (Version r2) [4]. To conduct imputation with the first two panels we utilized Michigan Imputation Server and to run imputation with the latter we used TOPMed Imputation Server [5]. Prior to imputation we conducted phasing with the software Eagle (version 2.4) [6]. After imputation, we compared the quality of imputed data based on the total number of SNPs with imputation quality statistic $R^2 > 0.7$. For the panel shown the best results we calculated a polygenic estimate of the height and body mass index (BMI) and collated them with the real height and BMI of the study participants as a proof of a high imputation quality. The polygenic scores were computed using PLINK v2 [7]. All the calculations were performed in R programming language (version 4.2.3).

Results: Imputation procedure allowed us to obtain 39131578 SNPs while using HRC reference panel, 47109535 SNPs when applying 1000 Genomes panel and

292136463 SNPs by utilizing TOPMed reference panel. After filtration of imputed data by $R^2 > 0.7$ the number of genetic variants decreased to 7358967, 7343241 and 7558595 for HRC, 1000 Genomes and TOPMed, respectively. The latter panel provided maximum number of high-quality imputed SNPs. Using the TOPMed-imputed genotypes we calculated the polygenic scores for height and BMI and compered them with the phenotypes from the study sample. The Pearson correlation coefficient for height and its estimate was equal to 0.35 (p-value = 3.8e–05) and the same coefficient for BMI and its polygenic score was 0.27 (p-value = 0.002).

Conclusion: TOPMed reference panel was considered as the most optimal for imputation of genotype data from the Russian sample.

Funding: This work was supported by the Russian Science Foundation (grant No. 22-15-20037) and the Government of the Novosibirsk region.

References

1. Leonova O.N. et al. A protocol for recruiting and analyzing the disease-oriented Russian disc degeneration study (RuDDS) biobank for functional omics studies of lumbar disc degeneration. *PloS One*. 2022;17(5):e0267384
2. Sudmant P.H. et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526(7571):75-81
3. McCarthy S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283
4. Taliun D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590(7845):290-299
5. Das S. et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287
6. Loh P.R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet*. 2016;48(11):1443-1448
7. Chang C.C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):s13742-015-0047-8

doi 10.18699/bgrs2024-4.3-07

Testing the potential of antihypertensive drugs for spinal pain treatment

Elgaeva E.E.[1, 2]*, Suri P.[3, 4, 5, 6], Williams F.M.K.[7], Freidin M.B.[8], Verzun D.A.[2], Tsepilov Y.A.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

[3] *Division of Rehabilitation Care Services, VA Puget Sound Health Care System, Seattle, WA, USA*

[4] *Seattle Epidemiologic Research and Information Center, VA Puget Sound Health Care System, Seattle, WA, USA*

[5] *Clinical Learning, Evidence, and Research (CLEAR) Center, University of Washington, Seattle, WA, USA*

[6] *Department of Rehabilitation Medicine, University of Washington, Seattle, WA, USA*

[7] *Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King's College London, London, UK*

[8] *School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK*

* *elizabeth.elgaeva@gmail.com*

Key words: Mendelian randomization; dorsalgia; beta-blockers; calcium channel blockers; angiotensin-converting enzyme inhibitors; statins

Motivation and Aim: Current treatment for spinal (neck and back) pain commonly turns out to be ineffective, making the search for new medications an important research goal. While developing a new drug takes years and substantial expense, the repurposing of existing medications can improve efficiency of this process. In this sense, testing the potential of antihypertensive drugs for spinal pain treatment may be promising, since the link between hypertension and back pain has been described elsewhere [1]. In this study, we aimed to assess the effect of four classes of cardiovascular medications (beta-blockers, calcium channel blockers, angiotensin-converting enzyme inhibitors, and statins) on spinal pain and answer the question of whether they can be useful for its treatment or prevention.

Methods and Algorithms: As the input data we used publicly available results of genome-wide association studies (GWAS) conducted in White Europeans for systolic blood pressure (N = 757,601) [2], low-density lipoprotein cholesterol (N = 173,082) [3], and spinal pain (N = 1,028,947) [4]. The first two were considered as exposures, and the latter was an outcome in Mendelian randomization (MR) analyses. All data passed quality control and unification in the GWAS-Map platform [5]. We followed the pipeline proposed by Gill and coauthors [6] to estimate the causal effect. In brief, we utilized the inverse-variance weighted MR approach [7], instrumental variables for beta-blockers, calcium channel blockers, and angiotensin-converting enzyme inhibitors preselected in resent work [6] and we found the instrumental variable for statins using PLINK v1.90b6.24 [8], the DrugBank [9], GeneCards and GeneHanser [10] databases. Effect of statin use on levels of low-density lipoprotein cholesterol was taken from the randomized clinical study [11]. For MR analyses we set the threshold for statistical significance at p-value < 0.0125 after correction for multiple testing. Additionally we estimated the detectable MR effect for each medication assuming 80 % statistical power.

Results: We observed no statistically significant effect on spinal pain for any of the four medication classes. However, point estimates for the MR analyses of beta-blockers were suggestively significant and demonstrated a protective effect on spinal pain (OR = 0.84 [95 % CI 0.72–0.98], *p*-value = 0.026). In contrast, point estimates for the MR analyses of calcium channel blockers suggested a detrimental effect on spinal pain (OR = 1.12 [95 % CI 1.02–1.24], *p*-value = 0.020) but this was not statistically significant after accounting for multiple testing. For other drugs MR results showed wide 95 % CIs (OR = 0.91 [95 % CI 0.49–1.71], *p*-value = 0.78 for and OR = 1.11 [95 % CI 0.96–1.27], *p*-value = 0.16 for statins) reflecting imprecise estimates and were not statistically significantly associated with spinal pain. In power calculations we estimated the detectable effects for beta-blockers, calcium channel blockers, angiotensin-converting enzyme inhibitors, and statins to be OR = 0.81, OR = 1.10, OR = 0.37, and OR = 0.80, respectively.

Conclusion: A protective effect of beta-blockers on spinal pain was suggested in the current study, consistent with findings from observational and bioinformatic studies of diverse pain phenotypes. The detrimental effect of calcium channel blockers on spinal pain identified in this work must be interpreted in the context of conflicting directions of effect on other pain traits in observational and bioinformatic studies.

Funding: This work was supported by the Russian Science Foundation (RSF) grant No. 22-15-20037 and the Government of the Novosibirsk region.

References

1. Kauppila L.I. Atherosclerosis and disc degeneration/low-back pain – a systematic review. *Eur J Vasc Endovasc Surg*. 2009;37(6):661-670. doi 10.1016/j.ejvs.2009.02.006
2. Evangelou E., Warren H.R., Mosen-Ansorena D., Mifsud B., Pazoki R. et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet*. 2018;50(10):1412-1425. doi 10.1038/s41588-018-0205-x
3. Willer C.J., Schmidt E.M., Sengupta S., Peloso G.M., Gustafsson S. et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274-1283. doi 10.1038/ng.2797
4. Bjornsdottir G., Stefansdottir L., Thorleifsson G., Sulem P., Norland K. et al. Rare SLC13A1 variants associate with intervertebral disc disorder highlighting role of sulfate in disc pathology. *Nat Commun*. 2022;13(1):634. doi 10.1038/s41467-022-28167-1
5. Shashkova T.I., Gorev D.D., Pakhomov E.D., Shadrina A.S., Sharapov S.Z., Tsepilov Y.A., Karssen L.C., Aulchenko Y.S. The GWAS-MAP platform for aggregation of results of genome-wide association studies and the GWAS-MAP|homo database of 70 billion genetic associations of human traits. *Vavilovskii Zhurnal Genet Selektsii*. 2020;24(8):876-884. doi 10.18699/VJ20.686 (in Russian)
6. Gill D., Georgakis M.K., Koskeridis F., Jiang L., Feng Q. et al. Use of genetic variants related to antihypertensive drugs to inform on efficacy and side effects. *Circulation*. 2019;140(4):270-279. doi 10.1161/Circulationaha.118.038814
7. Hemani G., Zheng J., Elsworth B., Wade K.H., Haberland V. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408. doi 10.7554/eLife.34408
8. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi 10.1086/519795
9. Wishart D.S., Knox C., Guo A.C., Shrivastava S., Hassanali M., Stothard P., Chang Z., Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006;34(Database issue):D668-D672. doi 10.1093/nar/gkj067
10. Fishilevich S., Nudel R., Rappaport N., Hadar R., Plaschkes I. et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*. 2017;2017:bax028. doi 10.1093/database/bax028
11. Cheung B.M., Lauder I.J., Lau C.P., Kumana C.R. Meta-analysis of large randomized controlled trials to evaluate the impact of statins on cardiovascular outcomes. *Br J Clin Pharmacol*. 2004;57(5):640-51. doi 10.1111/j.1365-2125.2003.02060.x

doi 10.18699/bgrs2024-4.3-08

The performance of machine learning approach in genome-wide association study of disease

Khvorykh G.[1]*, Belousov M.[2], Limborska S.[1], Khrunin A.[1]

[1] *National Research Centre "Kurchatov Institute", Moscow, Russia*

[2] *National Research University Higher School of Economics, Moscow, Russia*

** gennady.khvorykh@gmail.com*

Key words: genome-wide association study; single nucleotide polymorphism; machine learning; random forest; xgboos; shapley values

Motivation and Aim: The most common diseases like heart disease, strokes and diabetes, cancer, mental health, and autoimmune disorders are the most expensive and worried about in society. Individual differences in susceptibility to them are known to be substantially genetically determined. In spite of the tremendous development of data collection technologies in the field of molecular genetics occurring last decades, the dissection of genetic architecture of common diseases still remains a challenging task. Problems with sample sizes, environmental factors, pleiotropy are the main challenges in understanding common diseases [1]. Certain progress in this field is ascertained with machine learning (ML) algorithms. They may substitute traditional genome-wide association studies (GWAS) known for almost 20 years [2]. One ML-based approach consists of ranking single-nucleotide polymorphisms (SNPs) by their ability to predict case or control group the individual belongs to. The higher the rank, the more evident the association of SNP with a disease. However, the efficacy of this strategy demands both theoretical and empirical examinations. In this research, we investigated computationally the influence of data sizes on ranking of SNPs. Two metrics to rank the loci were considered: the feature impurity importance and SHapley Additive exPlanations (SHAP) [3]. The models were trained with Random Forest (RF) [4] and eXtreme Gradient Boosted trees (XGBoost) [5] using three data sets, each generated under the assumption of uncorrelated or correlated loci. The performance of the approach was characterized by accuracy, specificity and sensitivity. The results obtained were compared with those of classical GWAS. The observations thus made contributed to proper application of ML in associative studies.

Methods and Algorithms: Assuming alleles were uncorrelated, we generated genotypes for 25100 biallelic loci with Plink 1.9 tool (argument --simulate) [6]. The allele frequencies were uniformly distributed within the interval [0.05, 0.95]. One hundred loci were made to be associated with a disease. Another set of genotypes accounted for LD in groups of individuals of European ancestry. It was generated from haplotypes of CEU population (N = 95) from 1000 Genomes project [7] with Hapgen 2 [8] and included 23583 SNPs from chromosome 22. In this case, there were five SNPs associated with disease (rs4823464, rs137425, rs3761422, rs9606478, and rs7292279). The odds ratios (ORs) equaled to 1.5 and 2.25 for heterozygous and homozygous allele pairs, respectively. The disease loci were validated with Plink 1.9 tool (arguments --assoc). The case/control groups included 652/652, 4929/652, and 4929/4929 individuals, respectively. They are refereed further as small, unbalanced, and large datasets. The sizes of the unbalanced data matched those of the actual genotype-phenotype data that we had

previously used for GWAS of ischemic stroke [9]. The genotypes represented by symbols were converted into binary variables by one-hot encoding. The synthetic datasets thus obtained are available at https://github.com/inzilico/synthetic-data.git. The models were trained with RandomForestClassifier() from module scikit-learn 1.4 [10] and XGBClassifier() from package XGBoost 2.0.3 using the default arguments. The feature impurity importances were obtained from feature_importances_ attribute of trained models, the SHAP values were estimated with TreeExplainer() and Explainer() functions from shap package [11]. The performance of ML-based approach was assessed by accuracy, sensitivity, and specificity, where loci associated to disease were selected by percentiles varying from 95.0 till 99.9. The data were visualized with built-in functions from packages mentioned above and ggplot2 R package [12]. To automate the data processing, the custom scripts were written in Python 3.10 [13], R 4.2.1 [14], and GNU Bash 5.1.16 [15].

Results: The features sorted in descending way by feature impurity importance and SHAP values revealed the expected monotonic decrease in metrics with the increase in the rank for all datasets. However, the shape of the curves depended on the sample size and the type of metrics. The model fitted on unbalanced data did not predict the control samples (minor class). In the case of balanced datasets, the mean accuracies were around 0.5. Nevertheless, it allowed identifying the disease loci. However, the sensitivity estimated with 99 percentiles of SHAP values were 0.12, 0.27, and 0.60 for small, unbalanced and large datasets. The lists of selected loci for three datasets had less than 2.4 % of common items in pairwise comparisons. The distribution of neutral and disease loci in importance vs. SHAP space for three datasets (Fig. . **1**) demonstrated that it is not almost possible to subset the disease ones in the case of small dataset because of strong mixing them with the neutral ones. In the case of unbalanced dataset, the separation of a small portion of disease loci is possible. In the case of large dataset, an essential part of disease loci can be potentially separated, once the threshold values of metrics are estimated. The similar results were observed in the case of LD aware datasets.

Conclusion: The size and ratio of classes are crucial for the application of ML-based approach in associative studies. Its effective usage with ensemble algorithms

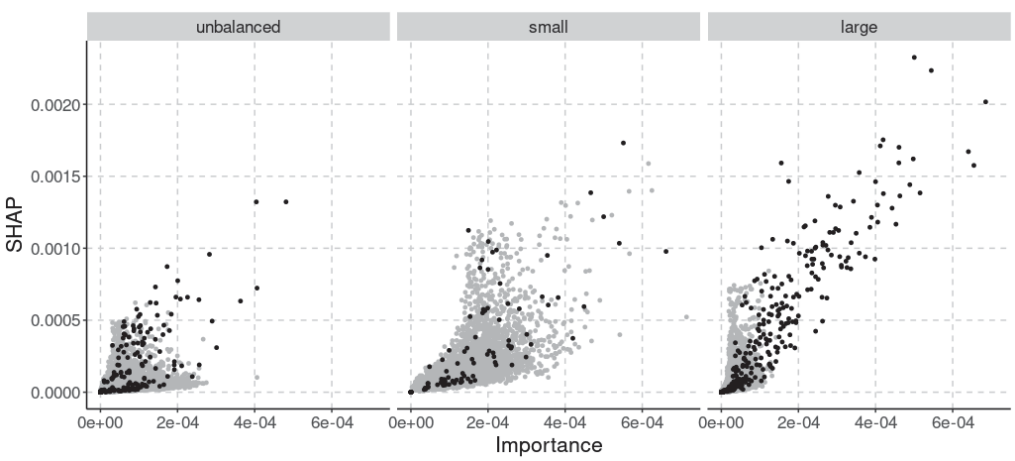


Fig. 1. The distribution of neutral (gray) and disease (black) loci in importance vs. SHAP space for three datasets without LD

(RF, XGBoost) would likely require balanced datasets with a total size of about 10000 individuals or more.

Funding: This research was supported by the Thematic plan of the National Research Centre "Kurchatov Institute" (5f.5.9., simulation of synthetic datasets) and by the Russian Science Foundation, grant number 23-14-00131 (fitting ML models and assessment of approach performance).

References

1. Visscher P.M. Challenges in understanding common disease. *Genome Med.* 2017;9:112. doi 10.1186/s13073-017-0506-1
2. Nicholls H.L., John C.R., Watson D.S., Munroe P.B., Barnes M.R., Cabrera C.P. Reaching the end-game for GWAS: machine learning approaches for the prioritization of complex disease loci. *Front Genet.* 2020;11:350. doi 10.3389/fgene.2020.00350
3. Lundberg S.M., Lee S.-I. A Unified approach to interpreting model predictions. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. 2017. [http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf]
4. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32. doi 10.1023/a:1010933404324
5. Chen T., Guestrin C. XGBoost: A Scalable Tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016. doi 10.1145/2939672.2939785
6. Chang C.C., Chow C.C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7. doi 10.1186/s13742-015-0047-8
7. Auton A., Abecasis G.R., Altshuler D.M., Durbin R.M., Bentley D.R. et al. A global reference for human genetic variation. *Nature*. 2015;526:68-74. doi 10.1038/nature15393
8. Su Z., Marchini J., Donnelly P. HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics*. 2011;27:2304-2305. doi 10.1093/bioinformatics/btr341
9. Khvorykh G.V., Sapozhnikov N.A., Limborska S.A., Khrunin A.V. Evaluation of density-based spatial clustering for identifying genomic loci associated with ischemic stroke in genome-wide data. *Int J Mol Sci*. 2023;24:15355. doi 10.3390/ijms242015355
10. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O. et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830
11. Lundberg S.M., Erion G., Chen H., DeGrave A., Prutkin J.M., Nair B. et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2:56-67. doi 10.1038/s42256-019-0138-9
12. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag, 2016
13. Van Rossum G., Drake F.L. Python 3 reference manual. Scotts Valley, CA: CreateSpace, 2009
14. R core team. R: a language and environment for statistical computing. 2021
15. GNU Bash (5.1) [Unix shell program]. 2020. https://www.gnu.org/software/bash/

doi 10.18699/bgrs2024-4.3-09

ELF5 SNPs and the risk of severe COVID-19 in Russian population

Loktionov A.[1, 2], Karpenko A.[1, 2], Kobzeva K.[2], Sergeeva V.[1], Bushueva O.[2, 3]*

[1] *Department of Anesthesia and Critical Care, Institute of Continuing Education, Kursk State Medical University, Kursk, Russia*

[2] *Laboratory of Genomic Research, Research Institute for Genetic and Molecular Epidemiology, Kursk State Medical University, Kursk, Russia*

[3] *Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, Kursk, Russia*

* *olga.bushueva@inbox.ru*

Key words: *ELF5*; SNPs; rs7949972; severe COVID-19

Motivation and Aim: COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), presents a spectrum of clinical manifestations, ranging from asymptomatic to severe organ failure. The variation in symptom severity poses a significant challenge for clinicians, underscoring the need to decipher the underlying factors contributing to disease severity. Addressing this complexity necessitates the identification of environmental and genetic risk factors. The aim of our study was to evaluate the impact of the polymorphic variants of *ELF5* gene rs7949972 and rs61882275, established by previously conducted GWAS, on the severity of COVID-19 in the population of Central Russia.

Methods and Algorithms: The study included 798 unrelated individuals from Central Russia, including 199 patients with COVID-19 hospitalized in the intensive care units of Kursk hospitals, and 599 control group patients with a mild form of COVID-19. Genotyping of rs7949972 and rs61882275 *ELF5* was carried out using real-time PCR according to a method developed in the laboratory of genomic research of the Research Institute for Genetic and Molecular Epidemiology. The thrombodynamics test was performed using the laboratory diagnostic system "Thrombodynamics Recorder TD-2". A log-additive regression model was used to assess associations. Bioinformatics resources were used to analyze the functional effects of SNPs.

Results: We determined that rs7949972 *ELF5* had a protective effect only in COVID-19 patients with a BMI < 30 (protective allele T, OR = 0.67, 95 % CI = 0.47–0.95, P = 0.02, P_{bonf} = 0.04). In this subgroup, we observed that allele T rs7949972 *ELF5* reduced the clot size at 30 min after coagulation activation (CS, μm, P = 0.02) and stationary spatial clot growth rates (Vst, μm/minutes, P = 0.02), and allele A rs61882275 *ELF5* was found to increase time of appearance of spontaneous clots (Tsp, P = 0.003).

ELF5, a member of the Ets transcription factor family, has been extensively investigated in breast cancer studies [1, 2]. Recent findings have underscored its relevance in COVID-19, revealing upregulation of key host factors (Ace2 and Tmprss4) in Elf5-overexpressing AT2 cells [3]. Moreover, ELF5, via cis-eQTL effects, modulates the expression of CAT, an antioxidant enzyme, in whole blood and the tibial artery, with elevated levels observed in COVID-19 patients [4]. Analysis of the impact of the risk allele C rs7949972 ELF5 on TFs binding sites suggests a potential role in exacerbating COVID-19 severity by positively regulating CD8-positive, alpha-beta T cell differentiation, negatively regulating CD4-positive, alpha-beta T cell differentiation, and

contributing to the defense response to viruses. Additionally, protective allele T rs7949972 is associated with improved lung function parameters, such as forced vital capacity (FVC), forced expired volume in 1 second (FEV1), FEV1 to FVC ratio, and peak expiratory flow, as evidenced by data from the Lung Knowledge Portal.

Conclusion: In the present study, we replicated associations of the SNP rs7949972 *ELF5* with severe COVID-19 within the Caucasian population of Central Russia, and determined BMI as modifier of the risk. We were the first to establish the impact of *ELF5* loci on thrombodynamic parameters such as CS, Vst and Tsp.

Funding: The study is supported by Kursk State Medical University.

References

1. Chakrabarti R., Hwang J., Andres Blanco M. et al. Elf5 inhibits the epithelial–mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. *Nat Cell Biol.* 2012;14(11):1212-1222. doi 10.1038/ncb2607
2. Kalyuga M., Gallego-Ortega D., Lee H.J. et al. ELF5 suppresses estrogen sensitivity and underpins the acquisition of antiestrogen resistance in luminal breast cancer. *PLoS Biol.* 2012;10(12):e1001461. doi 10.1371/journal.pbio.1001461
3. Pietzner M., Chua R.L., Wheeler E. et al. ELF5 is a potential respiratory epithelial cell-specific risk gene for severe COVID-19. *Nat Commun.* 2022;13(1):4484. doi 10.1038/s41467-022-31999-6
4. Martín-Fernández M., Aller R., Heredia-Rodríguez M. et al. Lipid peroxidation as a hallmark of severity in COVID-19 patients. *Redox Biol.* 2021;48:102181. doi 10.1016/j.redox.2021.102181

doi 10.18699/bgrs2024-4.3-10

Genetic dissection of spike productivity traits in spring barley

Rozanova I.V.[1]*, Grigoriev Y.N.[2], Kukoeva T.V.[2], Lukina K.A.[1], Shvachko N.A.[1], Kovaleva O.N.[1], Khlestkina E.K.[1]

[1] *N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *i.rozanova@vir.nw.ru*

Key words: barley; row; GWAS; SNP

Motivation and Aim: Barley (*Hordeum vulgare* L.) is one of the most commonly cultivated cereals worldwide. The total area occupied by this crop in Russia amounts to 8.8 million Ha. The main goal in barley breeding is to combine high yield with better grain quality. In Russia, the main areas are occupied by two-row barley. Six-row barley can be a carrier of many useful traits for the breeding process. The genes introgression that controls these traits to two-row varieties is a perspective task. When investigating the productivity traits related to the assessment of grain productivity, it is necessary to understand how much the loci that determine the spike row will affect this trait. The study of samples containing both two-row and six-row forms will allow us to identify new donors for breeding programs. An important task is to obtain the maximum possible yield with high grain quality.

Like other cultivated plants, barley requires constant improvement of varieties and adaptation to changing conditions. When breeding using traditional methods, the process of creating a variety includes many field tests and usually takes 12–15 years. Currently, selection methodology has been significantly enriched by the inclusion of genetic markers in the work. The most modern method is genome-wide association studies (GWAS), which is based on identifying SNPs associated with the desired trait. Over the past two decades, donor genetic diversity for various traits has been significantly expanded using GWAS.

The aim of this study is to identify loci associated with traits of barley grain productivity and to evaluate the influence of genes that control the number of rows in a spike on the studied phenotypes. The study examined traits that may affect productivity, such as “spike length”, “number of grains per main spike”, “weight of 1000 grains”, “grain weight of the main spike”. These characteristics are economically important because they are the main determinants of barley yield.

Methods and Algorithms: For candidate marker identification, the following analyses were carried out: phenotyping (structural), genotyping and association test. The genotyping was carried out using new chip Barley 50 K Illumina Infinium iSELECT. Using GWAS, significant SNPs associated with an agronomic trait were revealed over a three-year period. Then, we conducted association tests followed by meta-analysis of three years results. Based on the data obtained, a sample was created that included 129 only two-row barley. The sample was planted in four different locations, and the following was also carried out: phenotyping, genotyping and association analysis.

Results: In total, 64 SNPs (on 2H, 4H and 5H chromosomes) significantly associated with productivity traits were revealed in the sample studied, which consisted of two- and six-row barley varieties. No significant loci were identified when conducting association

analysis with row-type covariate. Accordingly, it turns out that all markers identified by GWAS will be correlated with genes that affect row-type. Next, a genome-wide association analysis was carried out only on 68 two-row varieties, which made it possible to revealed 42 SNPs on chromosomes 1H, 2H, 4H and 7H. This was confirmed by an association analysis of a sample consisting of 129 varieties of two-row barley at other locations. In both analyzes, both for only two-row barley and for combined 2/6-row barley, only the locus on chromosome 2H was common.

Conclusion: Our study represents a search for loci related to row type determination. It was found that genes controlling the rowing of the spike make a large contribution in the combined samples. Eleven chromosome regions were identified using GWAS. One of confirmed loci was previously described in the literature. The presence of a new chip of higher density allows more targeted detection of row-associated loci. We performed an experiment lasting for three consecutive years under different natural conditions and studied the effect of rowing on such agronomic traits as “spike length”, “number of grains per main spike”, “weight of 1000 grains”, and “grain weight of the main spike”.

Funding: The study is supported by VIR project No. 0481-2022-0007.

doi 10.18699/bgrs2024-4.3-13

Uncovering genetic factors of alcohol use disorder through GWAS and polygenic risk scores in East Slavs

Trofimov M.[1, 2*, #], Kudryavskiy V.[1, #], Shaheen L.[2, 3], Kovalenko E.[2], Vergasova K.[2], Kamelin A.[2], Rubinova V.[2], Kharitonov D.[2], Kim A.[2], Plotnikov N.[2], Elmuratov A.[2], Gainetdinov R.[4], Krupitsky E.[5, 6], Kibitov A.[5, 6], Rakitko A.[2, 7]

[1] *Skolkovo Institute of Science and Technology, Moscow, Russia*

[2] *Genotek Ltd., Moscow, Russia*

[3] *Moscow Institute of Physics and Technology, Dolgoprudny, Moscow region, Russia*

[4] *Institute of Translational Biomedicine, Saint Petersburg State University, St. Petersburg, Russia*

[5] *Saint Petersburg First Pavlov State Medical University, St. Petersburg, Russia*

[6] *V.M. Bekhterev National Medical Research Center for Psychiatry and Neurology, St. Petersburg, Russia*

[7] *Higher School of Economics, Faculty of Computer Science, Moscow, Russia*

* *m.trofimov@genotek.ru*

[#] *These authors contributed equally*

Key words: alcohol use disorder; polygenic risk score (PRS); East Slavic population; genome-wide association study (GWAS); genetic correlations; logistic regression

Motivation and Aim: Alcohol use disorder (AUD) significantly impacts over 200 health conditions, causing approximately 3 million annual deaths globally. Twin and adoption studies estimate that AUD is approximately 50 % heritable [1]. Multiple GWAS of alcohol-related traits, such as AUD [2] and alcohol dependence [3], have identified genetic associations primarily highlighting variants in genes directly related to alcohol metabolism (e. g., ADH1B and ALDH2). Our aims were to perform a GWAS of alcohol use disorder in the East Slavic population and develop a polygenic risk score (PRS). Additionally, we aimed to validate our findings in patients clinically diagnosed with AUD and in different ethnic groups represented in Russia.

Methods and Algorithms: We analyzed the genetic data of 41,575 individuals from Genotek Ltd. DNA was extracted and genotyped using Illumina Infinium Global Screening Array v.1-v.3 microarrays (650,000 SNPs). The Alcohol Use Disorder Identification Test (AUDIT), a ten-item questionnaire, was used to estimate alcohol drinking behavior and dependence. GWAS analyses were performed using PLINK v2.0 with the AUDIT total score as the dependent variable, accounting for age, sex, body mass index, and the first 20 principal components from genotype data as covariates. To calculate heritability and genetic correlations, we used LDSC tool v1.0.1 and summary statistics from UKB. PRS were generated using external GWAS summary statistics from P. Barr et al. [4], utilizing LDpred2 with default parameters. For comparison, a logistic regression model was built, classifying participants into ‘healthy’ and ‘case’ categories. Independent genotype data of patients diagnosed with alcohol dependence (AD) were acquired for PRS validation [5].

Results: From Genotek’s cohort, 41,575 individuals completed the AUDIT questionnaire. After filtering for the East Slavic population and removing close relatives, the sample size was 41,575 participants. In GWAS we detected one genome-wide significant SNP rs1229984 (odds ratio (OR) = 0.75, 95 % CI = 0.6–0.9) and p-value = 2.5×10^{-18} associated with AUD, located in the ADH1B gene. This SNP showed

significant associations in several ethnic groups, including Russians, Ukrainians, and Tatars, with consistent positive effect sizes. The estimated genetic heritability calculated from summary statistics of our GWAS was 6.4 %. Genetic correlations were assessed between AUD and 36 phenotypes, showing significant correlations with alcohol-related phenotypes, smoking status, and psychiatric disorders. PRS analysis using different AUDIT-T thresholds showed that a threshold of more than 14 performed slightly better, though the difference was not statistically significant. The OR for AUD were similar for both thresholds (Fig. 1). Our PRS demonstrated comparable performance to the external PRS, and it was validated in the independent clinical cohort of people diagnosed with alcohol dependence (AUC = 0.6, 95 % CI = 0.56–0.64). In addition, the PRS increased the predictive power of the multivariate model with nongenetic predictors by 1.33 % to AUC = 0.762 (p-value = 8×10^{-5}).

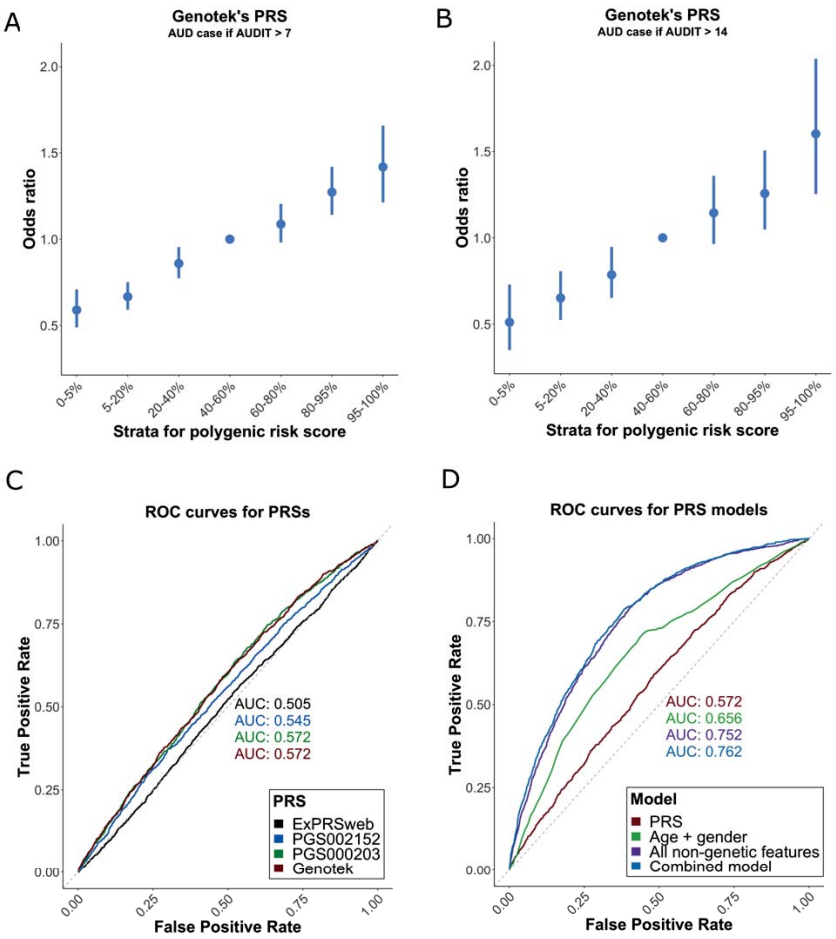


Fig. 1. PRS results. **A, B**: quantile plots of alcohol use disorder PRS built by LDpred-2. Based on PRS Individuals can be stratified into quantiles according to their risk score. It allows identifying groups of people with increased risk. Odds ratio represents comparison of PRS odds from different quantiles to the reference quantile (40–60 %). Bars represent confidence intervals. A, Individuals with AUDIT-T greater than 7 were defined as 'case'. B, Individuals with AUDIT-T greater than 14 were defined as 'case'. **C**, ROC curves plot for external PRSs and Genotek's PRS. **D**, ROC curves plot for models utilizing genetic and non-genetic features

Conclusion: We identified a well-known SNP in the ADH1B gene associated with AUD in the East Slavic population and found consistent associations across several ethnic groups. Our genetic correlation analysis confirmed expected relationships with alcohol- and smoking-related phenotypes. The performance of our PRS indicates that it can be applied across different European populations. The PRS was validated in an independent clinically established cohort, indicating that AUDIT-T is applicable for generating PRS to predict genetic predisposition to AUD. Further research may develop better scores for describing alcohol use disorder more comprehensively.

References

1. Hildebrand Karlén M. et al. Prevalence and heritability of alcohol use disorders in 18-year old Swedish twins. *Nordisk Alkohol Nark.* 2023;40:391-405
2. Kranzler H.R. et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat Commun.* 2019;10:1499
3. Walters R.K. et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci.* 2018;21:1656-1669
4. Barr P.B. et al. Using polygenic scores for identifying individuals at increased risk of substance use disorders in clinical and population samples. *Transl Psychiatry.* 2020;10:196
5. Levchenko A. et al. A Genome-Wide Association Study Reveals a BDNF-Centered Molecular Network Associated with Alcohol Dependence and Related Clinical Measures. *Biomedicines.* 2022;10:3007

5 Симпозиум «Эволюционная, популяционная геномика/генетика и молекулярная филогения: компьютерные и экспериментальные подходы»

Symposium “Evolutionary, population genomics/genetics and molecular phylogeny: computational and experimental approaches”

doi 10.18699/bgrs2024-5.1-01

Опсины гребневиков отряда Cydippida: филогения, структура и потенциальная возможность фоторецепции

Баяндина Ю.*, Кулешова О., Кривенко О.
ФИЦ «Институт биологии южный морей имени А.О. Ковалевского РАН», Севастополь, Россия
** sepulturka@mail.ru*

Ключевые слова*:* фоторецепция, гребневики, Cydippida, опсины, белковые последовательности.

Мотивация и цель. Опсины – ключевой элемент зрения и фоторецепции у большинства многоклеточных организмов. Они представляют собой G-белок связывающие трансмембранные рецепторы (GPCR), содержащие ретиналь в качестве светочувствительного хромофора. Эти фоторецепторные белки появились на ранних этапах эволюции многоклеточных животных (Metazoa). В ходе дальнейшей эволюции семейство опсинов претерпело значительную диверсификацию и расширение в различных таксонах животных. Ранние этапы эволюции опсинов остаются малоизученными [1]. Особый интерес представляет исследования фоторецепции у гребневиков, поскольку у этой древней группы многоклеточных животных сохранились предковые компоненты зрения [2, 3]. До недавнего времени знания об опсинах гребневиков были весьма ограниченными: были обнаружены только три паралога опсинов у двух видов гребневиков - *Mnemiopsis leidyi* и *Pleurobrachia bachei* [3]. В геномах и транскриптомах *Beroe* мы обнаружили три опсина, гомологичных опсинам *M. leidyi* и *P. bachei* [4]. Однако этих данных недостаточно для однозначного определения положения опсинов гребневиков на общем эволюционном дереве.

В данном исследовании мы провели анализ транскриптомных данных, доступных в базе NCBI, для различных видов гребневиков отряда Cydippida. одного из широко распространенных отрядов гребневиков. Целью анализа было выявление и идентификация белок-кодирующих последовательностей, которые могут быть связаны с восприятием света и фоторецепцией у этих морских животных.

Методы и алгоритмы. Мы проанализировали данные SRA для 24 транскриптомов Cydippida, найденных в NCBI: Cydippida sp. LM-2017 (SRR6074511), Cydippida sp. UF-2017 (SRR5892578), *Callianira antarctica* (SRR5892575), *Pleurobrachia pileus* (SRR6074514, SRR789901), *Pleurobrachia sp.* (SRR6074519, SRR6074520, SRR6074517, SRR5892573), *Pukia falcata* (SRR5892572*), Dryodora glandiformis* (SRR3048699, SRR777788), *Mertensiidae sp*. (SRR786492*), Euplokamis dunlapae* (SRR777663, SRR3048351), *Hormiphora californensis* (SRR10237136, SRR1992642), *Hormiphora palmata* (SRR6074513), *Lampea pancerina* (SRR3407163), *Lampea lacteal* (SRR3048349), *Bathyctena chuni* (SRR3048487), *Haeckelia rubra* (SRR3048520), *Haeckelia beehleri* (SRR3048894), *Aulacoctena acuminata* (SRR3049024), а также собственные данные транскриптома *Pleurobrachia pileus* (SRR26700624). Предварительную обработку данных, сборку транскриптома *de novo*, предсказание белковых последовательностей и автоматическую аннотацию проводили с помощью следующих программных средств: fastP v0.23.2, Trinity v2.13.2, TransDecoder v5.5.0, BLAST v.2.12.0, HMMER 3.2.1. SwisProt (11/17/2021) и Pfam-A (11/15/2021). Набор данных, полученный по запросу «Opsin» от SwisProt

и TrEMBL, был использован для поиска белков, участвующих в фоторецепции у Cydippida. Выравнивание последовательностей и филогенетический анализ мы проводили с помощью программ MAFFT v.7.48, IQ-TREE 1.6.12 с использованием набора последовательностей опсинов, представленных в работе Shettigar и соавторов в 2021 г. [5].

Результаты. Мы идентифицировали три гомолога белка семейства опсинов (обозначенных как Opsin1, 2, 3) в 13 из 16 проанализированных транскриптомов отдельных видов Cydippida. В транскриптоме Cydippida sp. UF-2017 обнаружен только гомолог Opsin1. В транскриптомах глубоководных видов *Bathyctena chuni* и *Aulacoctena acuminata*, обитающих на глубинах до 5000 метров, обнаружены только гомологи опсина третьей группы. Идентичность выровненных фрагментов опсинов Cydippida каждой группы, при сравнении с гомологичной последовательностью *Mnemiopsis leidyi* и *Pleurobrachia bachei*, по данным BlastP варьировала от 21 до 99% при покрытии от 47 до 99%.

На основе набора данных работы [5], дополненных последовательностями Beroida и Cydippida, мы реконструировали филогенетическое дерево опсинов животных с использованием метода максимального правдоподобия. Были идентифицированы две монофилетические группы опсинов гребневиков (или ктенопсинов).

Одна из монофилетических групп включает две сестринские клады: Ктенопсин1 и Ктенопсин2. По-видимому, они разошлись после дупликации гена опсина, присутствовавшего у общего предка Lobata, Cydippida и Beroida. В настоящее время нет свидетельств дальнейшего расширения семейства генов опсина в эволюции гребневиков. Согласно нашему филогенетическому анализу, группа, включающая ктенопсины 1 и 2, может быть сестринской по отношению ко всем другим опсинам многоклеточных животных, за исключением так называемых опсинов AnthozoanII. Ктенопсин3 образует монофилетическую кладу вне основных групп опсинов животных, которая могла первой отделиться от общего дерева эволюции этих белков у Metazoa.

Опсины Cydippida соответствуют стандартной 7-трансмембранной топологии GPCR. Анализ их структуры подтвердил особенности их структуры, выявленные ранее в опсинах *Beroe* и *Mnemiopsis*. В частности, все ктенопсины второго типа имеют большую вставку между пятой и шестой трансмембранными альфа-спиралями. В результате в области третьей цитоплазматической петли формируется две дополнительные спирали, отсутствующие в опсинах других животных. Эта область имеет решающее значение для взаимодействия между опсином и G-белком, и поэтому дополнительные цитоплазматические структуры могут влиять на особенности функционирования рецептора. Ктенопсины третьей группы имеют уникальную вставку из 13 аминокислотных остатков, следующих сразу за остатком Lys(K)296, через который ретиналь связывается с апопротеином у всех животных опсинов. Поскольку это может вызвать существенные изменения в структуре 7-ТМ и нарушать общую форму хромофорного кармана, способность Ктенопсинов3 связывать свет представляется мало вероятной.

Аминокислотные последовательности опсинов гребневиков сравнили с другими визуальными опсинами животных, чтобы оценить возможность их участия в фоторецепции. В Ктенопсине 1 и Ктенопсине 2 мы идентифицировали ключевые консервативные остатки и мотивы, необходимые для связывания с ретиналем и стабилизации конформационных состояний апопротеина. Уникальные консервативные мотивы были обнаружены в областях, обеспечивающих связь

опсина с G-белком. В последовательностях Ктенопсина3, отсутствовали консервативные остатки, отвечающие за поглощение фотона света в других животных опсинах и присутствовали некоторые консервативные мотивы, связанные со связыванием G-белка, характерные для других опсинов животных и отсутствующие в Ктенопсинах 1 и 2.
Заключение. Филогенетический анализ показал, что у гребневиков существует три типа опсинов. По-видимому, только два из них способны участвовать в фоторецепции, поскольку содержат все необходимые молекулярные компоненты для связывания с ретиналем и способность поддерживать конформационные изменения в различных световых условиях. Эти опсины образуют монофилетическую группу, и разошлись после дупликации гена опсина, общего для предков Lobata, Cydippida и Beroida. Нет никаких свидетельств дальнейшего расширения семейства генов опсина в эволюции гребневиков.
Финансирование: Работа выполнена в рамках Государственного задания ФИЦ ИнБЮМ «Биоразнообразие как основа устойчивого функционирования морских экосистем, критерии и научные принципы его сохранения» № 124022400148-4.

Ctenophore opsins from Cydippida: phylogeny, structure, and potential ability for photoreception

Baiandina Iu.*, Kuleshova O., Krivenko O.
A.O. Kovalevsky Institute of Biology of the Southern Seas of RAS, Sevastopol, Russia
** sepulturka@mail.ru*

Key words: photoreception, ctenophores, *Cydippida*, opsins, protein sequences.

Motivation and Aim: *Motivation and Aim*: Opsins form the basis of vision in most multicellular animals. Opsins are G-protein-coupled receptors (GPCRs) that contain a retinal chromophore. The opsin family originated early in animal evolution and expanded significantly across many different animal lineages. The early evolution of opsins remains poorly understood [1]. In this regard, light reception of ctenophores is important, as ancestral components of vision may have been conserved during the evolution of this ancient animal group [2, 3]. Until recently, knowledge of ctenophore opsins was limited to three paralogs in the *Mnemiopsis leidyi* and *Pleurobrachia bachei* [3]. Additionally, in the genomic and transcriptomic data from *Beroe*, we found three opsins that are homologous to those in *M. leidyi* and *P. bachei* [4]. However, the available data do not provide enough information to unambiguously determine the position of ctenophore opsins on the metazoan evolutionary tree or classify them into any known opsin group. In this study, we conducted an analysis of transcriptomic data available in the NCBI for various species of cydippid ctenophores, one of the most widespread orders of ctenophores. The aim of the analysis was to identify and characterize protein-coding sequences that may be associated with light perception and photoreception in these marine animals.
Methods and Algorithms We analyzed 24 samples of Cydippida SRA transcriptome data available at NCBI: Cydippida sp. LM-2017 (SRR6074511), Cydippida sp. UF-2017 (SRR5892578), *Callianira antarctica* (SRR5892575), *Pleurobrachia pileus*

(SRR6074514, SRR789901), *Pleurobrachia sp.* (SRR6074519, SRR6074520, SRR6074517, SRR5892573), *Pukia falcata* (SRR5892572*), Dryodora glandiformis* (SRR3048699, SRR777788), *Mertensiidae sp*. (SRR786492*), Euplokamis dunlapae* (SRR777663, SRR3048351), *Hormiphora californensis* (SRR10237136, SRR1992642), *Hormiphora palmata* (SRR6074513), *Lampea pancerina* (SRR3407163), *Lampea lacteal* (SRR3048349), *Bathyctena chuni* (SRR3048487), *Haeckelia rubra* (SRR3048520), *Haeckelia beehleri* (SRR3048894), *Aulacoctena acuminata* (SRR3049024), as well as own transcriptome data on *Pleurobrachia pileus* (SRR26700624). Data pre-preprocessing, *de novo* transcriptome assembling, prediction of protein sequences, and automatic annotation were carried out using the following programs: fastP v0.23.2, Trinity v2.13.2, TransDecoder v5.5.0, BLAST v.2.12.0, HMMER 3.2.1. SwisProt (11/17/2021) and Pfam-A (11/15/2021). The dataset established by the "Opsin" query from SwisProt and TrEMBL was used to search for proteins involved in photoreception in cydippids. We performed sequence alignment and phylogenetic analysis using programs MAFFT v.7.48, IQ-TREE 1.6.12 using a set of opsin sequences analyzed by Shettigar and co-authors in 2021 [5].

Results: We have identified three homologous opsin family proteins (noted as Opsin1, 2, 3) in the transcriptomes of 13 out of 16 species of Cydippida that were analyzed. In the transcriptome of Cydippida sp. UF-2017, only a homolog of Opsin1 was found. In the transcriptomes of *Bathyctena chuni* and *Aulacoctena acuminata*, which inhabit depths of 1,000 meters, homologs for Opsins 3 were only found. The identities of the aligned fragments of Opsin 1, 2 and 3 from Cydippida with the similar sequence of *Mnemiopsis leidyi* and *Pleurobrachia bachei*, according to BlastP, varied from 21% to 99%, with coverage ranging from 47 to 99%.

Based on the dataset from [5], expanded by Beroida and Cydippida sequences, we have reconstructed a phylogenetic tree of animal opsins using the maximum likelihood method. Two monophyletic groups of ctenophore opsins, or Ctenopsins, were identified. One of the monophyletic groups includes two sister clades: Ctenopsin1 and Ctenopsin2. These appear to have diverged after the duplication of an opsin gene shared by the common ancestor of Lobata, Cydippida, and Beroida. There is no evidence of further expansion of the opsin gene family during the evolution of ctenophores nowadays. According to our phylogenetic analysis, the Ctenopsins1-2 group may be a sister to all other metazoan opsins, except for the so-called AnthozoanII opsins, whose position on the phylogenetic tree of animal opsins needs clarification. Ctenopsin3 forms a monophyletic group that may be a basal clade outside the main animal groups.

Cydippida opsins conform to the standard 7-transmembrane topology of GPCRs. The analysis of their structure revealed specific features that had been identified in previous studies of opsins in *Beroe* and *Mnemiopsis*. Particularly, all Ctenopsin2 have a large insertion between the transmembrane alpha helices TM5 and TM6. It results in two extracytoplasmic helices, absent in other animal opsins. This region is crucial for the interaction between the opsin and the G-protein, and extra cytoplasmic structures may therefore influence the specificity of photoreceptor function. Ctenopsin3 homologues have a unique insertion of 13 amino acid residues just behind Lys(K)296, through which retinal binds to apoprotein in all animal opsins. It can cause substantial changes to the 7-TM structure and disrupt the overall shape of the chromophore pocket. This raises questions about the photoreceptor functionality of Ctenopsin3.

The amino acid sequences of three types of ctenophore opsins were compared to those of other animal opsins to assess the possibility of their involvement in photoreception.

In Ctenopsin 1 and Ctenopsin 2, we identified all the key conserved residues and motifs essential for retinal binding and the stabilization of conformational changes in the apoprotein. Unique conserved motifs were found in regions mediating opsin binding to the G-protein. In contrast, the Ctenopsin3 sequences lack conserved residues responsible for light absorption in other animal opsins. At the same time, there are some conserved motifs associated with G-protein binding that are characteristic of other animal opsins and absent in Ctenopsins 1 and 2.

Conclusion: A phylogenetic analysis revealed that there are three types of opsins shared among different species of ctenophores. Apparently, only two of them are capable of participating in photoreception, since they contain all the necessary molecular components for binding to retinal and the ability to maintain conformational changes under different light conditions. These opsins form a monophyletic group that diverged after the duplication of an opsin gene shared by the ancestors of Lobata, Cydippida, and Beroida. There is no evidence of further expansion of the opsin gene family during ctenophore evolution.

Funding: The study is supported by IBSS GA (No. 124022400148-4).

Список литературы/References

1. Hofmann K.P., Lamb T.D. Rhodopsin, light-sensor of vision. *Prog Retin Eye Res*. 2023;93:101116. doi 10.1016/j.preteyeres.2022.101116
2. Fleming J.F. et al. A novel approach to investigate the effect of tree reconstruction artifacts in single-gene analysis clarifies opsin evolution in nonbilaterian metazoans. *Genome Biol Evol*. 2020;12(2):3906-3916. doi 10.1093/gbe/evaa015
3. Schnitzler C.E. et al. Genomic organization, evolution, and expression of photoprotein and opsin genes in Mnemiopsis leidyi: a new view of ctenophore photocytes. *BMC Biol*. 2012;10:107. doi 10.1186/1741-7007-10-107
4. Baiandina I., Kuleshova O., Krivenko O. Light perception in Beroidae ctenophores: evidence from laboratory experiments and genomics data. In: Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022). Novosibirsk, 2022;29-30. doi 10.18699/SBB-2022-005
5. Shettigar N., Chakravarthy A., Umashankar S., Lakshmanan V., Palakodeti D., Gulyani A. Discovery of a body-wide photosensory array that matures in an adult-like animal and mediates eye-brain-independent movement and arousal. *Proc Natl Acad Sci USA*. 2021;118(20):e2021426118. doi 10.1073/pnas.2021426118

doi 10.18699/bgrs2024-5.1-02

Полногеномный анализ древней ДНК позволил уточнить филогеографию пещерных гиен в Северной Евразии

Боцманов Е.И.[1]*, Иванова А.О.[1], Крицкий А.А.[1], Павлова А.В.[1], Прокопьев Н.А.[1], Цедилина Т.Р.[1], Гимранов Д.О.[2], Хантемиров Д.Р.[3], Маликов Д.Г.[4], Климук Е.И.[1], Северинов К.В.[1]

[1] *ООО «Биотехнологический кампус», Москва, Россия*

[2] *Институт экологии растений и животных УрО РАН, Екатеринбург, Россия*

[3] *Уральский федеральный университет имени первого Президента России Б.Н. Ельцина, Екатеринбург, Россия*

[4] *Институт геологии и минералогии им. В.С. Соболева СО РАН, Новосибирск, Россия*

** ebotsmanov@biotc.ru*

Ключевые слова: палеогеномика; филогенетика; биоинформатика; гиены; древняя ДНК

Мотивация и цель: Пещерные гиены, населявшие большие области Евразийского континента вплоть до позднего плейстоцена, в силу значимых генетических и морфологических различий в литературе часто делятся на два вида: «европейский» (*Crocuta spelaea*) и «азиатский» (*C. ultima*) [1]. При этом взаимоотношение между видами рода *Crocuta* остается спорным и не до конца изученным. В частности, отсутствуют в достаточном количестве генетические данные для популяций, населявших центральную часть Северной Евразии. В настоящей работе были исследованы полные геномы двух особей рода *Crocuta* (ИЭРиЖ 3220/2242 и ИЭРиЖ 3220/2243), найденных в пещере Инейская (Республика Хакасия).

Методы и алгоритмы: Все работы с древней ДНК проводились в специальном «чистом» помещении на базе геномного центра «Биотехнологический Кампус» в соответствии со стандартными протоколами. Выделение древней ДНК было осуществлено в соответствии с протоколом Rohland et al., 2018 [2]. Для обеих особей были получены парноконцевые прочтения длинной 75 bp на платформе DNBSEQ-T7. Биоинформатический анализ проводили с использованием программ “FastP”, “bwa”, “angsd”. Аутентификация древней ДНК по специфическим паттернам повреждений осуществлялась при помощи “MapDamage2”. Для филогенетического анализа была использована программа “iqtree”. Для визуализации результатов использовали пакет seaborn (python) и itol. Среднее покрытие ядерного генома составило 2.19X и 0.81X для 3220/2242 и 3220/2243 соответственно.

Результаты: Филогенетический анализ митохондриальной ДНК позволил однозначно отнести обе особи к древней ветви гаплогруппы A (рис. 1). Этот результат косвенно подтверждает широкое распространение гаплогруппы A в центре Евразии, ранее показанное на примере пещерной гиены из Денисовой пещеры [3]. На основе новых и ранее опубликованных [4] данных полногеномного секвенирования проведен анализ главных компонент, позволяющий отнести обе особи из пещеры Инейская к виду *C. ultima* (рис. 2). Полученный результат соотносится с наблюдаемыми морфологическими отличиями [5].

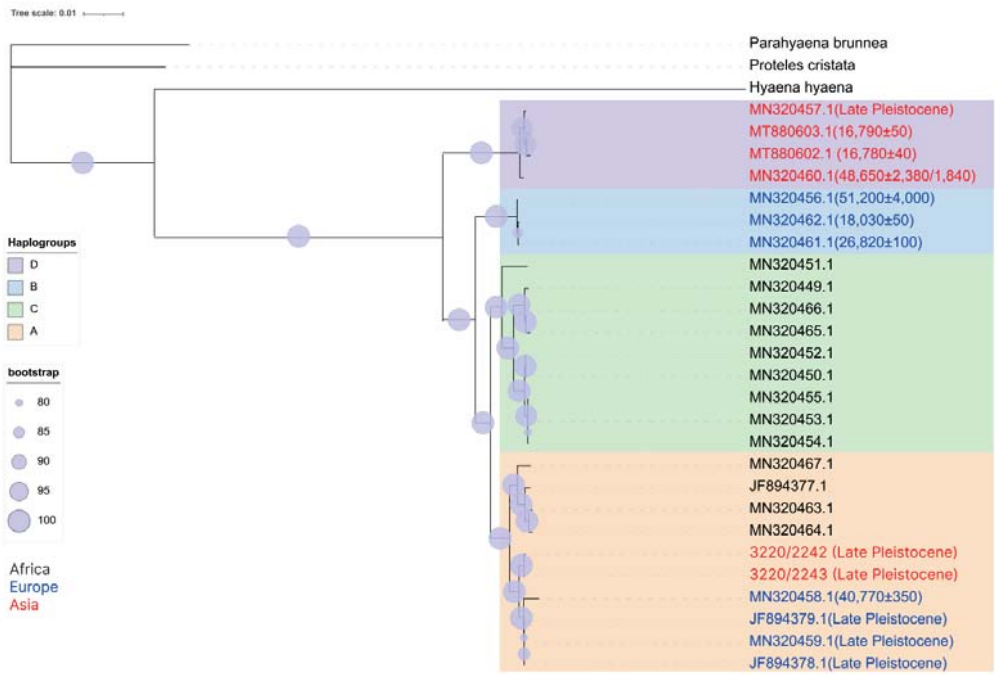


Рис. 1. Филогенетическое дерево представителей рода *Crocuta*, построенное на основе митохондриальных геномов. В качестве аутгруппы взяты *Parahyaena brunnea*, *Hyaena hyaena*, *Proteles cristatusa*. Синими кругами обозначены бутстрепы, определяющие статистическую достоверность дерева. Цветами клад обозначены ранее определенные гаплогруппы A, B, C, D [4, 6]. Текст выделен в соответствии с локацией исследованного образца. Для ископаемых образцов в скобках указана дата. Дерево построено методом максимального правдоподобия, модель замен – TPM2u+F+I+G4

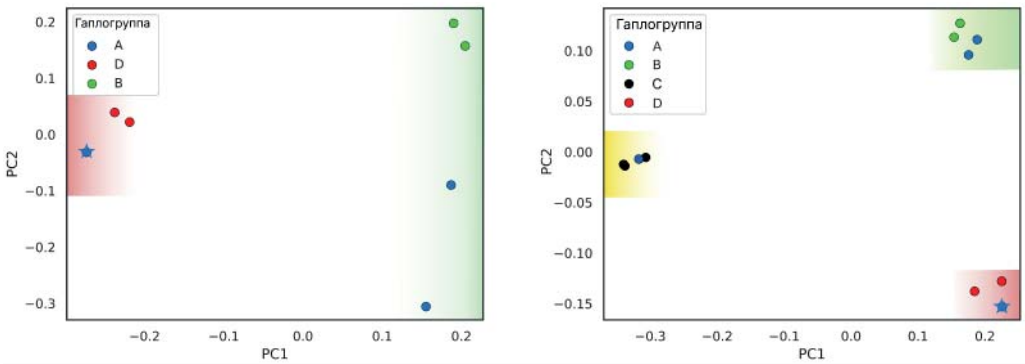


Рис. 2. Анализ главных компонент на основе ядерных однонуклеотидных полиморфизмов: слева – только для ископаемых образцов, справа – включая полные геномы современных пятнистых гиен. Особь 3220/2242 выделена звездочкой. Цвет точек соответствует ранее определенным митохондриальным гаплогруппам. Цвет закрашенных областей отражает принадлежность к следующим видам: желтый – современные *C. crocuta*, красный – *C. ultima*, зеленый – *C. spelaea*

Выводы: Результаты полногеномного секвенирования впервые подтверждают более широкое распространение вида *C. ultima* в центре Северной Евразии в позднем плейстоцене. Несоответствие результатов ядерного и митохондриального

филогенетического анализа может демонстрировать более тесные контакты *C. ultima* и *C. spelaea*, а также возможность продолжительного существования гибридных форм в зоне контакта их ареалов.
Финансирование: Исследование поддержано ООО «Биотехнологический кампус» и государственным заданием ИГМ СО РАН (№ 122041400243-9).

Whole genome sequencing of ancient DNA clarifies the phylogeography of cave hyenas in Northern Eurasia

Botsmanov E.[1]*, Ivanova A.[1], Kritsky A.[1], Pavlova A.[1], Prokopev N.[1], Tsedilina T.[1], Gimranov D.[2], Khantemirov D.[3], Malikov D.[4], Klimuk E.[1], Severinov K.[1]

[1] *"Biotech Campus" LLC, Moscow, Russia*

[2] *Institute of Plant and Animal Ecology, UB RAS, Ekaterinburg, Russia*

[3] *Ural Federal University named after the First President of Russia B.N. Yeltsin, Ekaterinburg, Russia*

[4] *V.S. Sobolev Institute of Geology and Mineralogy, SB RAS, Novosibirsk, Russia*

** ebotsmanov@biotc.ru*

Key words: paleogenomics; phylogenetics; bioinformatics; hyenas; ancient DNA

Motivation and Aim: Cave hyenas, which inhabited large areas of the Eurasian continent until the late Pleistocene, are often divided into two species, European (*Crocuta spelaea*) and Asian (*C. ultima*), due to significant genetic and morphological differences [1]. The relationship between these species remains controversial and not fully understood. In particular, sufficient genetic data are lacking for populations that inhabited the central part of Northern Eurasia. In this work, the whole genomes of two animals of the genus *Crocuta* (IPAE 3220/2242 and IPAE 3220/2243) found in the Ineyskaya cave (Republic of Khakassia) were investigated.
Methods and Algorithms: All work with ancient DNA was carried out in a clean room facility at the Biotech Campus genomic center in accordance with standard protocols. The isolation of ancient DNA was performed according to the protocol of Rohland et al., 2018 [2]. For both samples, DNA libraries were prepared according to MGI protocol, and 75 bp paired-end reads were obtained on the DNBSEQ-T7 platform. Bioinformatics analysis was performed using the software tools "FastP", "bwa", and "angsd". Authentication of ancient DNA by specific damage patterns was carried out using the "MapDamage2" package. The "iqtree" tool was used for phylogenetic analysis. The package "seaborn" (python) and "itol" were applied for the results visualization. The nuclear genome mean coverage was 2.19X and 0.81X for 3220/2242 and 3220/2243 respectively.
Results: Phylogenetic analysis of mitochondrial DNA allowed unambiguously assign both individuals to the ancient branch of haplogroup A (Fig. 1). The same haplogroup was observed for a cave hyena from Denisova Cave [3]. Thus, haplogroup A was widely distributed the center of Eurasia. On the basis of new and previously published whole-genome sequencing data [4], we performed a principal component analysis, which allowed both animals from the Ineyskaya Cave to be classified as *C. ultima* (Fig. 2), which correlates with observed morphological data [5].

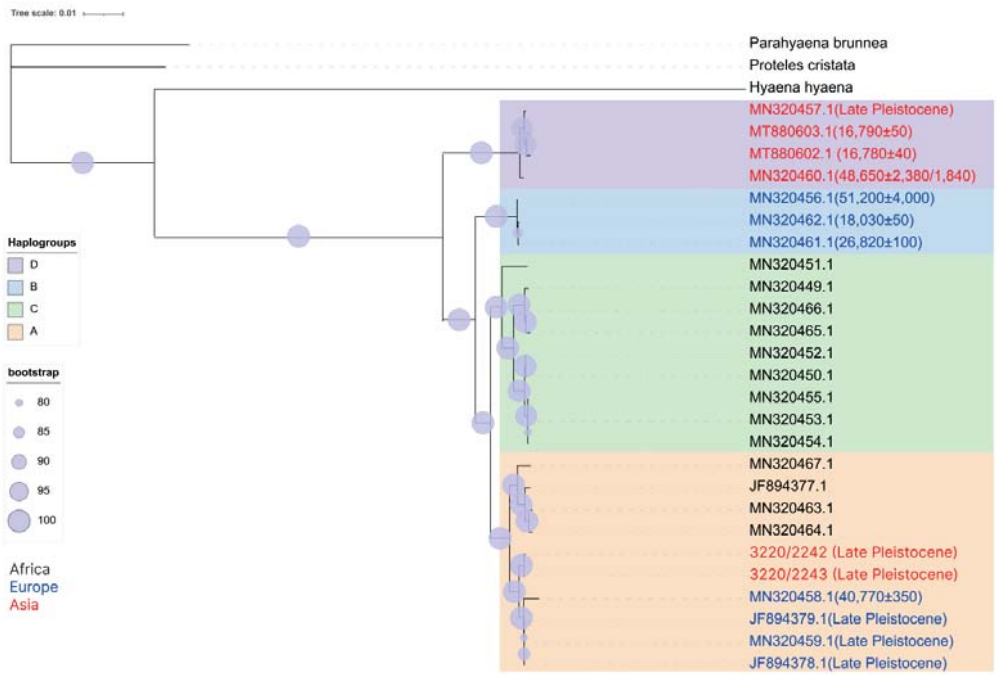


Fig. 1. Phylogenetic tree of representatives of the genus *Crocuta* based on mitochondrial genomes. *Parahyaena brunnea*, *Hyaena hyaena*, *Proteles cristatusa* are taken as outgroups. Blue circles indicate bootstraps defining the statistical confidence of the tree. Clade colors indicate previously defined haplogroups: A, B, C, D [4, 6]. Text is highlighted according to the location of the studied sample. For fossil specimens, the date is indicated in parentheses. The tree was constructed by the maximum likelihood method, the substitution model is TPM2u+F+I+G4

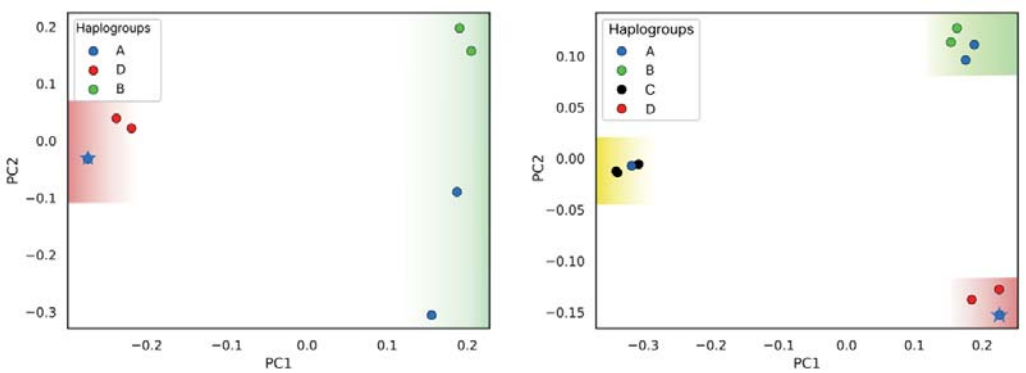


Fig. 2. Principal component analysis based on nuclear single nucleotide polymorphisms. The figure on the left is for fossil specimens only, the figure on the right includes the complete genomes of modern spotted hyenas. Individual 3220/2242 is marked as a star symbol. The color of the dots corresponds to previously defined mitochondrial haplogroups. The color of the shaded areas reflects the following species: Yellow – modern *C. crocuta*, Red – *C. ultima*, Green – *C. spelaea*

Conclusion: The results of whole-genome sequencing show wide distribution of *C. ultima* species in the center of Northern Eurasia during the Late Pleistocene. The discrepancy between the results of nuclear and mitochondrial phylogenetic analysis may

indicate closer contacts between *C. ultima* and *C. spelaea*, as well as a possibility of prolonged existence of hybrid forms in the contact zone of their ranges.
Funding: The study was supported by Biotech Campus LLC and the state assignment of IGM SB RAS (No. 122041400243-9).

Список литературы/References

1. Lewis M., Werdelin L. A revision of the genus Crocuta (Mammalia, Hyaenidae). *Palaeontogr Abt A Palazool Stratigr*. 2022;322:1-115. doi 10.1127/pala/2022/0120
2. Rohland N., Glocke I., Aximu-Petri A., Meyer M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat Protoc*. 2018;13(11):2447-2461. doi 10.1038/s41596-018-0050-5
3. Rohland N., Pollack J.L., Nagel D., Beauval C., Airvaux J., Pääbo S., Hofreiter M. The population history of extant and extinct hyenas. *Mol Biol Evol*. 2005;22(12):2435-2443. doi 10.1093/molbev/msi244
4. Westbury M.V., Hartmann S., Barlow A. et al., Hyena paleogenomes reveal a complex evolutionary history of cross-continental gene flow between spotted and cave hyena. *Sci Adv*. 2020;6:eaay0456. doi 10.1126/sciadv.aay0456
5. Хантемиров Д.Р., Маликов Д.Г., Гимранов Д.О. Пещерная гиена Crocuta cf. ultima из пещеры Инейская (Хакасия). В: Закономерности эволюции и биостратиграфия. Материалы LXX сессии Палеонтологического общества РАН. СПб.: Картфабрика ин-та Карпинского, 2024
6. Sheng G.L, Soubrier J., Liu J.Y. et al. Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, Crocuta crocuta. *Mol Ecol*. 2014;23(3):522-533. doi 10.1111/mec.12576

doi 10.18699/bgrs2024-5.1-03

Регуляторная геномика агрессивного поведения на неканонической модели лисы *Vulpes vulpes*

Дудко Н.[1]*, Андреева Т.[1, 2, 3], Манахов А.[1, 2, 3], Шепелева Д.[4], Шихевич С.[4], Харламова А.[4], Рогаев Е.[1, 3, 5]

[1] *Научный центр генетики и наук о жизни, Научно-технологический университет «Сириус», Федеральная территория «Сириус», Россия*

[2] *Лаборатория эволюционной геномики, Отдел геномики и генетики человека, Институт общей генетики имени Н.И. Вавилова РАН, Москва, Россия*

[3] *Центр генетики и генетических технологий, биологический факультет, МГУ имени М.В. Ломоносова, Москва, Россия*

[4] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[5] *Медицинская школа Чан Массачусетского университета, департамент психиатрии, Шрусбери, США*

* *dudko@rogaevlab.ru*

Ключевые слова: агрессия; Vulpes vulpes; эпигенетика; регуляция

Поведение – сложный фенотип, регулируемый различными сигнальными путями нейросетей, включая серотониновую, окситоцинергическую и гипоталамо-гипофизарную. Для изучения генетических и эпигенетических факторов, определяющих агрессивное поведение по отношению к человеку, была использована модель лис Vulpes vulpes. Были глубоко секвенированы геномы агрессивных, дружелюбных и неселектированных лисиц, выявлены однонуклеотидные полиморфизмы и структурные варианты, встречающиеся с разной частотой в исследуемых группах, также обнаружены различия в ретрокопиях некоторых генов. Были получены и проанализированы эпигенетические профили (h3k4me3 и h3k27ac) фронтальной коры лис, демонстрирующих разное поведение по отношению к человеку. Между агрессивными и дружелюбными животными обнаружены значимые изменения в уровнях обогащения по гистоновым меткам h3k4me3 и h3k27ac генов, входящих в сигнальные пути эстрогенового рецептора, «ERK (MAPK/ERK)», «нейроактивного взаимодействия лиганда и рецептора», «аксонального поиска пути» и «серотонинергического синапса».

Regulatory genomics of aggressive behavior on the non-canonical model of fox *Vulpes vulpes*

Dudko N.[1]*, Andreeva T.[1, 2], Manakhov A.[1, 2, 3], Shepeleva D.[4], Shihevich S.[4], Kharlamova A.[4], Rogaev E.[1, 5]

[1] *Center of Genetics and Life Sciences, Federal Territory Sirius, Sirius University of Science and Technology, Sochi, Russia*

[2] *Laboratory of Evolutionary Genomics, Department of Genomics and Human Genetics, Vavilov Institute of General Genetics, RAS, Moscow, Russia*

[3] *Center for Genetics and Genetic Technologies, Moscow State University, Moscow, Russia*
[2] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*
[4] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[5] *Department of Psychiatry, Umass Chan Medical School, Shrewsbury, MA, USA*
* *dudko@rogaevlab.ru*

Key words: aggression; fox; Vulpes vulpes; epigenetics; regulation

A complex phenotype, behavior depends on a variety of signaling pathways including oxytocinergic, serotoninergic, and hypothalamic-pituitary ones. The Vulpes vulpes model was used to determine the genetic components that influence aggressive behavior toward humans. SNPs and SVs that distinguished across groups were found in the genomes of aggressive, tame, and non-selected animals. There were found to be several genes with retrocopies lacking in one group and present in another. For the first time, Vulpes vulpes frontal brain epigenetic profiles were acquired, examined, and contrasted between the groups under study. Significant differences were observed in the levels of enrichment for histone marks h3k4me3 and h3k27ac of genes participating in the signaling pathways of the estrogen receptor, "ERK (MAPK/ERK)", "neuroactive interaction of ligand and receptor", "axonal pathfinding", and "serotonergic" synapse" between aggressive and tame animals.

doi 10.18699/bgrs2024-5.1-04

Отсутствие связи между цитотипом лабораторных линий *Drosophila melanogaster* при внутривидовом PM гибридном дисгенезе и *P* мобильным элементом

Захаренко Л.[1]*, Ромашева Е.[2], Илинский Ю.[1]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный педагогический университет, Новосибирск, Россия*

* *zakharlp@bionet.nsc.ru*

Ключевые слова: *Drosophila*; цитотип; гибридный дисгенез; P-элемент

Мотивация и цель: Внутривидовой гибридный дисгенез (ГД) у *Drosophila melanogaster* проявляется в атрофии гонад для некоторых пар линий в одном из направлений скрещивания [1]. В учебники вошло представление о том, что PM ГД обусловлен массовым перемещение *P* мобильного элемента. Однако ряд фактов заставляет усомниться в верности этой гипотезы: низкая скорость перемещения мобильных генетических элементов (МГЭ); ГД проявляется при повышенной температуре содержания дисгенного потомства, однако транскрипция МГЭ в этом диапазоне температур не меняется; часто отсутствует асимметрия в содержании *P*-элемента в родительских линиях; при прочих равных условиях линии, способные индуцировать симптомы ГД, уникальны [2–4]. Только референсные для PM ГД линии дают хорошо воспроизводимый результат в одном из направлений скрещивания в виде 100 % атрофии гонад у дисгенных самок при повышенной температуре содержания (29 °C).

В данной работе мы проверяли стабильность цитотипа (описание цитотипов в табл. 1) в линиях, где атрофия гонад проявлялась менее чем в 100 % случаев, и как связано проявление симптомов ГД с содержанием *P*-элемента в геноме.

Методы и алгоритмы: Для оценки цитотипов тестируемых линий проводили скрещивания с референсными для PM ГД линиями: самок тестируемых линий скрещивали с самцами линии Harwich, а самцов тестируемых линий – с самками линии Canton-S. Результаты тестирования сравнивали с данными, полученными 10 лет назад на этих же линиях [5]. Наличие *P*-элемента в геноме тестируемых линий проверяли с помощью ПЦР-анализа.

Результаты: Ранее в 2014 г. был проведен анализ цитотипов 52 линий *D. melanogaster* из четырех природных популяций: 14 линий из популяции Athens, Georgia, USA (GA), 17 линий из популяции Montpelier, France (FR), 13 линий из популяции Accra, Ghana (GH) [1]. Несмотря на содержание полноразмерной копии *P*-элемента в геноме, линии распределились по цитотипам следующим образом: доминировал Q-цитотип, не реагирующий на скрещивания с референсными для PM ГД линиями (31 линия), 12 линий имели P-цитотип, который индуцировал атрофию гонад, как правило, не более чем в 50 % случаев; 6 линий имели реактивный M’-цитотип, и 3 линии – P’-цитотип, вызывающий атрофию гонад в обоих направлениях скрещивания [1]. В данной работе в анализ были взяты по три линии каждого цитотипа: P, P’, Q и M’ (см. табл. 1). За 10 лет в 7 из 12 исследованных линий *D. melanogaster* цитотип сменился по сравнению с данными,

полученными в 2014 г. Линии GH-1 и GH-9 сохранили нейтральный Q цитотип, линия GA-16 сохранила P'-цитотип, реагирующий на скрещивания в обоих направлениях, GH-18 сохранила индуцирующий ГД P-цитотип, FR-29 – M' реактивный цитотип. Среди оставшихся линий доминирует нейтральный Q-цитотип. Несмотря на длительное содержание линий в лаборатории, все линии сохранили в геноме полноразмерную копию *P*-элемента (рис. 1).

Таблица 1. Характеристика цитотипов тестируемых линий и сравнение с данными 2023 г.

Характеристика цитотипов исследованных (t) линий	2014	2023	Сохранили цитотип
P атрофия гонад >10 % при скрещивании t-самцов с самками Canton-S	3	1	1
P' атрофия гонад >10 % в обоих направлениях скрещивания	3	1	1
M (нет P-элементов в геноме) или **M'** (есть P-элементы) Атрофия гонад >10 % при скрещивании t-самок с самцами Harwich	3	2	1
Q Нет атрофии гонад в обоих направлениях скрещивания	3	8	2

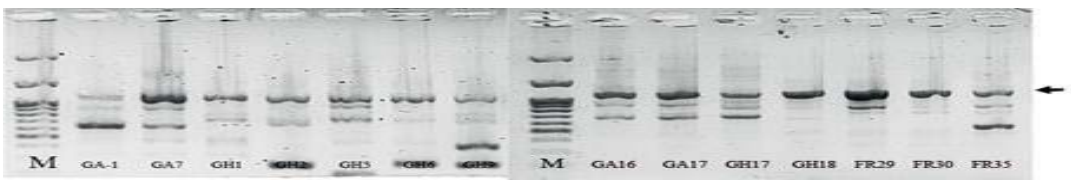


Рис. 1. ПЦР-анализ на содержание *P*-элемента в геноме тестируемых линий. Стрелкой указано положение центральной части P-элемента

Выводы:

1. Сравнение полученных результатов с данными 2014 г. показало, что только 42 % линий сохранили прежний цитотип: в двух линиях сохранился Q-цитотип, по одной линии сохранили P', P и M' цитотипы, в остальных линиях доминирует Q-цитотип, т. е. характеристики природных линий, как проявляющих, так и не проявляющих дисгенные свойства, нестабильны.
2. ПЦР анализ показал, что длительное содержание линий в лабораторных условиях не привело к утрате полноразмерного *P*-элемента в тестируемых линиях. Потеря способности индуцировать дисгенез P-линиями GA-6 и GA-7 не сопровождается потерей *P*-элемента. Результаты работы дают основание предполагать, что внутривидовой PM гибридный дисгенез вряд ли связан с *P*-элементом. Причиной нестабильности цитотипа может быть полиморфизм в тестируемых линиях, который накапливается и/или проявляется через десятки поколений содержания линий в лаборатории.

Финансирование: Исследование поддержано бюджетным проектом FWNR-2022-0019).

The cytotype instability of the *Drosophila melanogaster* lines does not depend on the *P*-element during intraspecific hybrid dysgenesis

Zakharenko L.[1]*, Romasheva E.[2], Ilinsky Yu.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State Pedagogical University, Novosibirsk, Russia*

* *zakharlp@bionet.nsc.ru*

Key words: *Drosophila*; cytotype; hybrid dysgenesis; P-element

Motivation and Aim: Intraspecific hybrid dysgenesis (HD) manifests itself in gonadal atrophy for some pairs of Drosophila melanogaster lines in one of the cross-directions [1]. The textbooks include the idea that the PM HD is caused by the massive movement of *P* transposable element (TE). However, a number of facts cast doubt on the validity of this hypothesis: low speed of movement of TEs; HD manifests itself only at elevated temperature of keeping dysgenic offspring, however, TE transcription does not change in this temperature range; often there is no asymmetry in the content of the *P*-element in the parental lines; all other things being equal, lines capable of inducing symptoms of HD are unique [2–4]. Only reference for PM HD lines give a well-reproducible result in one of the cross-directions as 100 % gonadal atrophy in dysgenic females at elevated temperature (29 °C).

In this work, we checked the stability of the cytotype (description of cytotypes in Table 1) in lines where gonadal atrophy manifested itself in less than 100 % of cases, and how the manifestation of HD symptoms is related to the presence of the *P*-element in the genome.

Methods and Algorithms: To evaluate the cytotypes of the tested lines, crosses were carried out with reference for PM HD lines: females of the tested lines were crossed with males of the Harwich line, and males of the tested lines with females of the Canton-S line. The test results were compared with data obtained 10 years ago on the same lines [5]. The presence of the P-element in the genome of the tested lines was checked using PCR analysis.

Results: Earlier in 2014, an analysis of the cytotypes of 52 lines of *D. melanogaster* from four natural populations was carried out: 14 lines from the population of Athens, Georgia, USA (GA), 17 lines from the population of Montpelier, France (FR), 13 lines from the population of Accra, Ghana (GH) [1]. Despite the content of a full-length copy of the *P*-element in the genome, the lines were distributed by cytotype as follows: the Q-cytotype dominated, which did not respond to crossings with reference for PM HD lines (31 lines), 12 lines had a P-cytotype, which induced gonadal atrophy no more than in 50 % of cases; 6 lines had a reactive M'-cytotype, and 3 lines had a P'-cytotype, causing gonadal atrophy in both cross-directions [1].

In this work, three lines of each cytotype were taken into analysis: P, P', Q and M' (Table 1). Over 10 years, in 7 of the 12 tested *D. melanogaster* lines, the cytotype changed compared to the data obtained in 2014. Lines GH-1 and GH-9 retained the neutral Q cytotype, line GA-16 retained the P'-cytotype, which is responsive to crossings in both directions, GH-18 retained the HD-inducing P cytotype, FR-29 retained the M' reactive cytotype. Among the remaining lines, the neutral Q cytotype dominates. Despite

the long-term maintenance of the lines in the laboratory, all lines retained a full-length copy of the *P*-element in the genome (Fig. 1).

Table 1. Characteristics of the cytotypes of the tested lines and comparison with the data of 2023

Characteristics of cytotypes of the tested (t) lines	2014	2023	Saved cytotype
P gonadal atrophy >10 % when crossing t-males with Canton-S females	3	1	1
P' gonadal atrophy >10 % in both directions of crossing	3	1	1
M (no *P*-elements in the genome) or M' (there are *P*-elements) gonadal atrophy >10% when crossing t-females with Harwich males	3	2	1
Q No gonadal atrophy in both directions of crossing	3	8	2

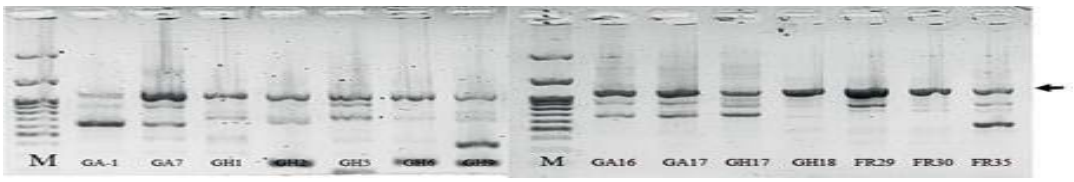


Fig. 1. PCR analysis for the presence of the *P*-element in the genome of the tested lines. The arrow indicates the position of the central part of the *P*-element

Conclusions: 1. Comparison of the results obtained with the data of 2014 showed that only 42 % of the lines retained the same cytotype: in two lines the Q-cytotype was preserved, one line retained the P', P or M' cytotypes, in the remaining lines the Q-cytotype dominates. Thus, the characteristics of natural lines, both exhibiting and not exhibiting dysgenic properties, are unstable.

2. PCR analysis showed that long-term maintenance of the lines in laboratory conditions did not lead to the loss of the full-length *P*-element in the tested lines. The loss of the ability to induce hybrid dysgenesis by P-lines GA-6 and GA-7 is not accompanied by loss of the *P*-element. The results of the work suggest that intraspecific PM hybrid dysgenesis is unlikely to be associated with the *P*-element. The cause of cytotype instability may be a polymorphism in the tested lines that accumulates and/or manifests itself through dozens of generations of line maintenance in the laboratory.

Funding: The study is supported by budget project #FWNR-2022-0019).

Список литературы/References

1. Kidwell M.G. Hybrid dysgenesis in Drosophila melanogaster: nature and inheritance of P element regulation. *Genetics.* 1985;111:337-350. doi 10.1093/genetics/111.2.337
2. Itoh M., Sasai N., Inoue Y., Watada M. P elements and P-M characteristics in natural populations of Drosophila melanogaster in the southern most islands of Japan and in Taiwan. *Heredity (Edinb.).* 2001;86(Pt. 2):206-212. doi 10.1046/j.1365-2540.2001.00817.x
3. Mombach D.M., da Fontoura Gomes T.M.F., Loreto E.L.S. Stress does not induce a general transcription of transposable elements in *Drosophila*. *Mol Biol Rep*. 2022;49:9033-9040. doi 10.1007/s11033-022-07839-7
4. Paulouskaya O., Romero-Soriano V., Ramirez-Lanzas C., Price T.A.R., Betancourt A.J. Levels of P-element-induced hybrid dysgenesis in Drosophila simulans are uncorrelated with levels of P-element piRNAs. *G3 (Bethesda).* 2023;13(2):jkac324. doi 10.1093/g3journal/jkac324
5. Ignatenko O.M., Zakharenko L.P., Dorogova N.V., Fedorova S.A. P elements and the determinants of hybrid dysgenesis have different dynamics of propagation in *Drosophila melanogaster* populations. *Genetica.* 2015;143:751-759. doi 10.1007/s10709-015-9872-z

doi 10.18699/bgrs2024-5.1-05

Особенности вторичной структуры спейсерного участка ITS2 рРНК трематод надсемейства Echinostomatoidea

Калинина К.А.[1, 2]*, Татонова Ю.В.[1, 2], Щелканов М.Ю.[1, 2]

[1] *ФГБНУ «НИИ эпидемиологии и микробиологии им. Г.П. Сомова» Роспотребнадзора, Владивосток, Россия*

[2] *Федеральный научный центр биоразнообразия наземной биоты Восточной Азии ДВО РАН, Владивосток, Россия*

*zzzzssssqqq@yandex.ru

Ключевые слова: Echinostomatoidea; *Psilotrema*; *Sphaeridiotrema*; *Echinochasmus*; ITS2; филогенетика; вторичные структуры

Мотивация и цель: Echinostomatoidea – это большая космополитическая группа дигенеев, которые имеют сложный жизненный цикл с одним или несколькими промежуточными и окончательными хозяевами. Промежуточными хозяевами служат различные виды моллюсков, рыбы и амфибии [1–3]. Подавляющее большинство трематод этой группы в качестве окончательного хозяина используют птиц, а также млекопитающих, рептилий и рыб [1, 2, 4]. Некоторые виды были зарегистрированы у человека, что повышает эпидемиологическую значимость данной группы трематод [5] и интерес к изучению разных аспектов выживаемости и адаптации этих паразитов. Одно из таких направлений – это изучение транскрибируемых спейсерных регионов. Они являются важным компонентом процессинга рРНК и биогенеза активных рибосомных субъединиц, в связи с этим получение и анализ вторичных структур этих участков востребованы для понимания биологической роли этих молекул [6]. Эта информация, в свою очередь, может быть полезна для понимания формирования адаптаций разных групп паразитов. Поэтому целью данной работы стало изучение вторичных структур участка ITS2 рРНК для трематод надсемейства Echinostomatoidea.

Методы и алгоритмы: Для бионформатического анализа были использованы виды *Echinochasmus pseudobeleocephalus* [7], *Sphaeridiotrema ussuriensis*, *S. aziaticus*, *S. pyriforme* и *Psilotrema limosum* [8]. Выделение ДНК проводили методом HotSHOT [9]. Далее все образцы были амплифицированы и секвенированы по Сэнгеру. Выравнивание нуклеотидных последовательностей произведено с помощью двух методов: ClustalW [10] и T-coffee [11]. Филогенетические конструкции построены с использованием байесовского вывода (BI) в программе MrBayes 3.1.2 [12] и алгоритма максимального правдоподобия (ML) в программе PhyML 3.1 [13]. Для построения вторичных структур спейсерного участка ITS2 использованы веб-серверы: UNAfold, RNAfold и NUPACK. Для сравнительного анализа были также изучены нуклеотидные последовательности других представители надсемейства из базы данных NCBI (GenBank).

Результаты: Филогенетический анализ на основе последовательностей участка ITS2 рДНК показал, что положение всех семейств на реконструкциях совпадает с данными, полученными Ткачом и соавторами по гену 28S рРНК [4]. Узлы имеют

относительно высокие значения статистических поддержек, а генетические дистанции имеют широкий диапазон межродовых отличий (47–61 %).
Согласно ранее опубликованным данным, пространственное строение участка ITS2 рРНК для большинства трематод имеет типичную четырехдоменную структуру [14]. Но мы обнаружили, что в некоторых случаях структура может быть вариабельна. Это зависит от длины последовательности: чем она длиннее, тем больше петельных доменов присутствует в смоделированной структуре. На разработанной нами общей модели вторичной структуры для надсемейства Echinostomatoidea наиболее консервативными являются первый и второй домены.
Выводы: Несмотря на то что маркер ITS2 рДНК во многих группах трематод не подходит для проведения филогенетического анализа, например, в надсемействе Opisthorchioidea [15], он дает хорошие результаты при оценке филогенетических отношений внутри надсемейства Echinostomatoidea, как было показано ранее для семейств Echinostomatidae [16] и Echinochasmidae [17]. В настоящей работе филогенетические реконструкции по участку ITS2 рДНК позволили уточнить таксономический статус некоторых представителей надсемейства Echinostomatoidea, а морфологические особенности вторичной структуры рРНК подтвердили возможность использования маркера для дальнейшего изучения этой группы трематод.
Финансирование: Исследования выполнены в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № 121031000154-4) и государственного задания № 123022200035-0 «Структура природных очагов паразитарных заболеваний на юге Дальнего Востока России».

Features of the secondary structure of the ITS2 rRNA spacer region of trematodes from the superfamily Echinostomatoidea

Kalinina K.A.[1, 2]*, Tatonova Y.V.[1, 2], Shchelkanov M.Y.[1, 2]

[1] *G.P. Somov Research Institute of Epidemiology and Microbiology, Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing, Vladivostok, Russia*

[2] *Federal Scientific Center for Biodiversity of Terrestrial Biota of East Asia, FEB RAS, Vladivostok, Russia*

* *zzzzssssqqq@yandex.ru*

Key words: Echinostomatoidea; *Psilotrema*; *Sphaeridiotrema*; *Echinochasmus*; ITS2; phylogenetics; secondary structures

Motivation and Aim: The Echinostomatoidea are a large cosmopolitan group of digeneans that have a complicated life cycle with one or more intermediate and definitive hosts. Intermediate hosts serve as various species of mollusks, fish and amphibians [1–3]. The vast majority of trematodes in this group use birds, as well as mammals, reptiles and fish as the definitive host [1, 2, 4]. Some species have been recorded in humans, which increase the epidemiological significance of this group of trematodes [5] and interest in studying different aspects of the survival and adaptation of these parasites. One of such areas is the study of transcribed spacer regions. They are very important for rRNA processing and the biogenesis of active ribosomal subunits; therefore, obtaining and analyzing the secondary structures of these regions is in demand for understanding the biological role of the analyzed molecules [6]. Based on mentioned above, the aim of

this work is to analyze the features of secondary structures of the ITS2 rRNA region for trematodes of the superfamily Echinostomatoidea.

Methods and Algorithms: The species *Echinochasmus pseudobeleocephalus* [7], *Sphaeridiotrema ussuriensis*, and *S. aziaticus*, *S. pyriforme* and *Psilotrema limosum* [8] were used for bioinformatics analysis. The HotSHOT method [9] was used for DNA extraction. Further, all the samples were amplified and sequenced. Alignment of the nucleotide sequences was performed using ClustalW [10] and T-coffee [11]. Phylogenetic trees were constructed using Bayesian inference (BI) in MrBayes 3.1.2 [12] and maximum likelihood (ML) in PhyML 3.1 [13]. Secondary structures of the ITS2 spacer region were created using the following web servers: UNAfold, RNAfold, and NUPACK. For comparative analysis, the nucleotide sequences of other representatives of the superfamily from the GenBank were also studied.

Results: Phylogenetic analysis based on the nucleotide sequences of ITS2 rRNA region showed that the position of all families in the obtained reconstructions is similar to data obtained by Tkach and co-authors based on the 28S rRNA gene [4]. Nodes have relatively high statistical support values, and genetic distances have a wide range of intergeneric differences (47–61 %).

According to previously published data, the spatial structure of the ITS2 rRNA region has a typical four-domain structure for most trematodes [14]. But we found that the structure is variable in some cases. This variability depends on the length of the sequence: the longer it is the more loop domains are present in the modeled structure. Based on the general secondary structure model developed for the superfamily Echinostomatoidea, the first and second domains are the most conservative.

Conclusion: Although the ITS2 rDNA marker is not suitable for phylogenetic analysis in many trematode groups, for example, in the superfamily Opisthorchioidea [15], it presented good results in assessing phylogenetic relationships within the superfamily Echinostomatoidea, as previously shown for the families Echinostomatidae [16] and Echinochasmidae [17]. In this study, the obtained phylogenetic reconstructions using ITS2 rDNA made it possible to clarify the taxonomic status of some representatives of the superfamily Echinostomatoidea, and the morphological features of the secondary structure of rRNA confirmed the possibility of using the marker for further study of this group of trematodes.

Funding: The research was carried out within the framework of the state assignment from the Ministry of Science and Higher Education of the Russian Federation (topic No. 121031000154-4) and government assignment no. 123022200035-0 “The structure of natural foci of parasitic diseases in the south of the Russian Far East”.

Список литературы/References

1. Догель В.А. Зоология беспозвоночных. Москва: Высшая школа, 1981
[Dogel V.A. Invertebrate zoology. Moscow: Vysshaya shkola, 1981 (in Russian)]
2. Галактионов К.В. Эволюция и биологическая радиации трематод: краткий очерк идей и мнений. Гл. II. *Труды Зоологического института РАН*. 2016;320(S4):74-126
[Galaktionov K.V. Evolution and biological radiation of trematodes: a brief outline of ideas and opinions. Chapter II. *Trudy Zoologicheskogo Instituta RAN*. 2016;320(S4):74-126 (in Russian)]
3. Щелканов Е.М., Тишина Е.А., Мануков Ю.И., Сапрыкин В.П. Биотрансформация ксенобиотиков моллюсками (Mollusca L., 1758) – индикаторами состояния водных экосистем. *Географическая среда и живые системы*. 2024(3)
[Shchelkanov E.M., Tishina E.A., Manukov Y.I., Saprykin V.P. Biotransformation of xenobiotics by mollusks (Mollusca L., 1758) – indicators of the state of aquatic ecosystems. *Geograficheskaya sreda i zhivyye sistemy* 2024(3) (in Russian)]

4. Tkach V.V., Kudlai O., Kostadinova A. Molecular phylogeny and systematics of the Echinostomatoidea Looss, 1899 (Platyhelminthes: Digenea). *Int J Parasitol*. 2016;46(3):171-185
5. Sayasone S., Tesana S., Utzinger J., Hatz C., Akkhavong K., Odermatt P. Rare human infection with the trematode Echinochasmus japonicus in Lao PDR. *Parasitol Int*. 2020;5(1):106-109
6. Tatonova Y.V., Chelomina G.N., Besprosvannykh V.V. Genetic diversity of nuclear ITS1–5.8S–ITS2 rDNA sequence in *Clonorchis sinensis* Cobbold, 1875 (Trematoda: Opisthorchidae) from the Russian Far East. *Parasitol Int*. 2012;61:664-674
7. Kalinina K.A., Besprozvannykha V.V., Tatonova Y.V., Shchelkanov M.Y. A Description of *Echinochasmus pseudobeleocephalus* n. sp. (Echinochasmidae) based on morphological and molecular data. *Animals (Basel)*. 2023;13(20):3236
8. Kalinina K.A., Tatonova Y.V., Besprozvannykha V.V. New species of *Psilotrema* and *Sphaeridiotrema* (Psilostomidae Odhner, 1913) in the east Asian region: Morphology of developmental stages and genetic data. *Parasitol Int*. 2022;88:102-554
9. Truett G.E., Heege P., Mynatt R.L., Truett A.A., Walker J.A., Warman M.L. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques*. 2000;29:52-54
10. Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947-2948
11. Notredame C., Higgins D.G., Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment 1. *J Mol Biol*. 2000;302(1):205-217
12. Ronquist F., Huelsenbeck J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572-1574
13. Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307-321
14. Morgan A.T., Blair J.D. Trematode and monogenean rRNA ITS2 secondary structures support a four-domain model. *J Mol Evol*. 1998;47:406-419
15. Thaenkhama U., Blairb D., Nawaa Y., Waikagula J. Families Opisthorchiidae and Heterophyidae: Are they distinct? *Parasitol Int*. 2012;61(1)90-93
16. Izrailskaia A.V., Besprozvannykh V.V., Tatonova Y.V. *Echinostoma chankensis* nom. nov., other *Echinostoma* spp. and *Isthmiophora hortensis* in East Asia: Morphology, molecular data and phylogeny within Echinostomatidae. *Parasitology*. 2021;148(11):1366-1382
17. Tatonova Y.V., Izrailskaia A.V., Besprozvannykh V.V. *Stephanoprora amurensis* sp. nov., *Echinochasmus milvi* Yamaguti, 1939 and E. suifunensis Besprozvannykh, 1991 from the Russian southern Far East and their phylogenetic relationships within the Echinochasmidae Odhner 1910. *Parasitology*. 2020;147(13):1469-1479

doi 10.18699/bgrs2024-5.1-06

Анализ представленности семейств транспозонов в геномах домашней и дикой формы северного оленя *Rangifer tarandus*

Коноров Е.[1, 2]*, Семина М.[1], Онохов А.[1], Лайшев К.[1, 3]

[1] *Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия*

[2] *Федеральный научный центр пищевых систем им. В.М. Горбатова РАН, Москва, Россия*

[3] *Санкт-Петербургский федеральный исследовательский центр РАН, Санкт-Петербург, Россия*

* *konorov@vigg.ru*

Ключевые слова: *Rangifer tarandus*; LINE ретротранспозоны; доместикация

Мотивация и цель: На сегодняшний день доступно несколько сборок геномов северного оленя в открытых базах данных, в том числе одна на уровне хромосом [1–3]. В данных работах рассматривались специфичные для *Rangifer tarandus* гены, в том числе быстро меняющиеся и относящиеся к перепредставленным генным онтологиям. Около 38 % генома составляли повторяющиеся последовательности, но различия между особями из разных популяций не были предметом геномных исследований. Фракция повторяющихся последовательностей в геноме может разительно отличаться между подвидами одного вида копытных животных [4], при этом данные различия могут приводит к адаптации [5] и коррелировать с хозяйственно значимыми признаками. В данной работе мы сравнили два генома диких и два генома домашних северных оленей по распространенности повторяющихся элементов и, в частности, по числу копий LINE ретротранспозонов.

Методы и алгоритмы: С помощью Illumina NovaSeq секвенированы полные геномы двух домашних и двух диких особей северного оленя для последующей сборки и анализа повторяющихся последовательностей для характеризации различий между дикой и домашней морфами. Глубина секвенирования была выше 30x для каждого из геномов. Проведена очистка геномных прочтений для каждого образца с помощью trimmomatic [6], при этом были удалены геномные прочтения с низким качеством, адаптеры и другие артефакты секвенирования, а также произведен поиск контаминации. Произведен процесс сборки генома с помощью SPAdes [7]. Поиск повторяющихся последовательностей осуществлялся с помощью RepeatMasker и hmmer.

Результаты: N50 полученных сборок составляли 10.3 и 9.7 kB для домашних северных оленей и 11.7 и 12.2 kB для диких оленей, GC% состав не отличался и составлял 41.55 %.

Отдельно было изучено число копий BovB LINE элементов и L1 LINE. У домашних особей было обнаружено 101 и 97 копий BovB (DF000280309.1 RanTar-1.51) в геноме, в то время как в геномах диких оленей было найдено 120 и 127 полных BovB LINE. L1 LINE, специфичный для *R. tarandus* (DF000280306.1 RanTar-1.226) был найден в примерно одинаковом числе копий для всех четырех особей (121 и 125 для домашних, 125 и 128 для диких). По другим видоспецифичным семействам L1 (DF000280305, DF000280315 и DF000280316) также не обнаружено значительных различий.

Вместе с этим можно отметить, что различия внутри семейства ретротранспозонов у диких оленей (по p-дистанциям) в среднем выше, чем различия между таковыми копиями у домашних оленей. Это согласуется с неопубликованными данными по тем же выборкам, отсеквенированным с меньшим покрытием, но в большем объеме, что как ядерный, так и митохондриальный геном диких оленей обладает большей изменчивостью по сравнению с домашней морфой.
Выводы: Нами были аннотированы геномы повторяющихся последовательностей двух диких оленей из Магаданской области и двух домашних оленей из Мурманской области. Дикие олени обладают большим количеством копий отдельных семейств BovB LINE ретротранспозонов, а также большей вариабельностью между копиями. Поскольку, как показано в других работах по доместикации и локальным адаптациям у млекопитающих, LINE могут менять паттерн изменчивости для последующего отбора, в дальнейшем планируется расширить материал на нескольких популяциях с более низкой глубиной секвенирования
Финансирование: Работа выполнена при финансовой поддержке гранта РНФ № 22-16-00062.

Analysis of the transposon families abundance in the genomes of domestic and wild reindeer *Rangifer tarandus*

Konorov E.A.[1, 2]*, Semina M.T.[1], Onokhov A.A.[1], Layshev K.A.[1, 3]
[1] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*
[2] *Gorbatov Federal Research Center for Food Systems, RAS, Moscow, Russia*
[3] *St. Petersburg Federal Research Center, RAS, St. Petersburg, Russia*
* *konorov@vigg.ru*

Key words: *Rangifer tarandus*; LINE retrotransposons; domestication

Motivation and Aim: To date, several genome assemblies of reindeer *R. tarandus* are available in open databases, including one at the chromosome level [1–3]. These studies examined species-specific genes, including those which are rapidly changing and with overrepresented gene ontologies. About 38 % of the reindeer genome was repetitive sequences, but differences between individuals from different populations have not been the subject of genomic studies. The fraction of repeated sequences in the genome can differ drastically between subspecies of the same species of ungulates [4], and these differences can lead to adaptation [5] and correlate with economically significant traits of domesticated animals. In present work, we compared two genomes of wild and two genomes of domestic reindeer on prevalence of repetitive elements and, in particular, the copy number of LINE retrotransposons.
Methods and Algorithms: Using Illumina NovaSeq, the complete genomes of two domestic and two wild reindeer were sequenced and assembled Then we made repeat sequence analysis to characterize differences between wild and domestic morphs. The sequencing depth was above 30x for each of the genomes. The genomic reads for each sample were cleaned using trimmomatic [6], low-quality genomic reads, adapters and other sequencing artifacts were removed. The genomes was de novo assembled using SPAdes [7]. The search for repeated sequences was carried out using RepeatMasker and hmmer.

Results: The N50 of the resulting assemblies were 10.3 and 9.7 kB for domestic reindeer and 11.7 and 12.2 kB for wild deer, the GC% composition did not differ and was 41.55 %.

The copy numbers of BovB LINE elements and L1 LINE were studied separately. Domestic deer had 101 and 97 copies of BovB (DF000280309.1 RanTar-1.51) in the genome, while 120 and 127 complete BovB LINEs were found in wild deer genomes. L1 LINE specific for *R. tarandus* (DF000280306.1 RanTar-1.226) was found in approximately the same number of copies for all 4 individuals (121 and 125 for domestic, 125 and 128 for wild). No significant differences were also found for other species-specific L1 families (DF000280305, DF000280315 and DF000280316).

It should be noted that the differences within the family of retrotransposons in wild deer (according to p-distances) are on average higher than the differences between such copies in domestic deer. This is consistent with unpublished data from the same samples sequenced with less coverage, but in a larger volume, that both the nuclear and mitochondrial genomes of wild deer have greater variability compared to the domestic morph.

Conclusion: We have annotated repeat sequences in genomes of two wild deer from the Magadan region and two domestic deer from the Murmansk region. Wild deer have higher copy numbers of BovB LINE retrotransposon from several families and also have greater intercopy variability. As has been shown in other work on domestication and local adaptation in mammals LINEs can change the pattern of variation for subsequent selection, we are planning to expand the material on several populations with lower sequencing depth in the future.

Funding: The study is supported by Russian Science Foundation grant (No. 22-16-00062).

Список литературы/References

1. Taylor R.S., Horn R.L., Zhang X., Golding G.B., Manseau M., Wilson P.J. The caribou (*Rangifer tarandus*) genome. *Genes*. 2019;10(7):540
2. Poisson W., Prunier J., Carrier A., Gilbert I., Mastromonaco G., Albert V., Robert C. Chromosome-level assembly of the *Rangifer tarandus* genome and validation of cervid and bovid evolution insights. *BMC Genomics*. 2023;24(1):142
3. Weldenegodguad M., Pokharel K., Ming Y., Honkatukia M., Peippo J., Reilas T., Kantanen J. Genome sequence and comparative analysis of reindeer (*Rangifer tarandus*) in northern Eurasia. *Sci Rep.* 2020;10(1):8980
4. Petersen M., Winter S., Coimbra R., de Jong M.J., Kapitonov V.V., Nilsson M.A. Population analysis of retrotransposons in giraffe genomes supports RTE decline and widespread LINE1 activity in Giraffidae. *Mobile DNA.* 2021;12,1-13
5. Pan Z., Li S., Liu Q., Wang Z., Zhou Z., Di R., Li Y. Rapid evolution of a retro-transposable hotspot of ovine genome underlies the alteration of BMP2 expression and development of fat tails. *BMC Genomics.* 2019;20:261
6. Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120
7. Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Computat Biol.* 2012;19(5):455-477

doi 10.18699/bgrs2024-5.1-07

Полиморфизм ряда генов-кандидатов экономически значимых количественных признаков у крупного рогатого скота костромской породы и поиск его связи с продуктивными признаками

Лазебная И.В.[1]*, Лазебный О.Е.[2]

[1] *Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия*

[2] *Федеральное государственное бюджетное учреждение науки Институт биологии развития им. Н.К. Кольцова РАН, Москва, Россия*

** Lazebnaya@mail.ru*

Ключевые слова: крупный рогатый скот; гены-кандидаты; признаки продуктивности

Мотивация и цель: Селекция сельскохозяйственных животных, основанная на использовании ассоциации аллелей определенных генов с экономически значимыми количественными и качественными признаками (отбор с помощью маркеров – Marker-Assisted Selection, MAS), является одним из эффективных методов достижения высоких результатов в аграрном производстве. Поиск и исследование спектров генов, относящихся к потенциальным физиологическим и позиционным ДНК-маркерам, проводится с конца прошлого столетия [1]. Достигнуты значимые результаты. Определены спектры генов, продукты которых вовлечены в формирование определенных количественных признаков молочной и мясной продуктивности крупного рогатого скота [2].

Несмотря на значительные результаты, достигнутые в направлении поиска новых и получения данных для уже известных SNP-маркеров, остается не совсем ясным, насколько универсален характер установленных ассоциаций и возможно ли их использование в селекции всех пород независимо от направления продуктивности этих пород, их происхождения и достижений ранее проведенной селекции. При этом имеют значение и частотные характеристики распределения наиболее ценных аллелей во внутрипородных выборках крупного рогатого скота, которые должны быть оценены, как и генотипирование быков-производителей, до проведения соответствующего отбора.

Целью настоящего исследования являлось изучение генетической структуры в выборках костромской породы крупного рогатого скота обоих полов с использованием SNP ряда генов-кандидатов (*GH*, *GHR*, *DGAT*1, *PRL*, *SCD*, *RORC*, *Lep*, *LepR*, *CSN*3) и выявление их ассоциаций с признаками молочной продуктивности дочерей исследованных быков.

Методы и алгоритмы: Генотипировали быков (N = 75) и коров (N = 120) костромской породы по SNP ряда генов-кандидатов: гормона роста (*GH*, L127V) и его рецептора (*GHR*, S555G), диацилглицерол ацилтрансферазы (*DGAT*1, K232A), пролактина (*PRL*, g.35108342A>G) и каппа-казеина (*CSN3*, Ile136Thr и Ala148Asp), стеароил-КоА-десатуразы (*SCD*, A293V) С-рецептора ретиноевой кислоты (*RORC*, g.3984A>G), стеароил-КоА-десатуразы (*SCD*, A293V), лептина

(*Lep*, Y7F) и его рецептора *LepR*, T945M), как описано [3–6]. Популяционно-генетический анализ проводили с использованием программ GenAlEx v.6.5 и STRUCTURE 2.3.4. Ассоциативный анализ признаков молочной продуктивности (удой (кг), содержание (%) и выход жира (кг)) дочерей общей численностью более 950 выполняли в связи с генотипами их отцов, быков-производителей, в программе STATISTICA 10.

Результаты: В исследованных выборках не выявлено значимого отклонения распределения наблюдаемых частот генотипов от их распределения согласно Харди–Вайнбергу. Установлена дифференциация выборок друг от друга по маркерам генов *CSN*3, *DGAT*I и *LepR* согласно приведенным значениям вероятности для значений G_{ST} (табл. 1) и с учетом поправки на множественные сравнения Бенджамини–Хохберга [7].

Таблица 1. Значения коэффициента G_{ST} и их вероятность

Locus	*CSN3*	*PRL*	*GH*	*DGATI*	*SCD*	*RORC(A)*	*GHR*	*Lep*	*LepR*	Tot
Gst	0.016	–0.002	–0.002	0.044	0.007	–0.002	–0.002	0.006	0.034	0.011
P	0.006	0.748	0.627	0.001	0.039	0.957	0.663	0.067	0.001	0.001

Уровень дифференциации на основе комплекса маркеров с использованием индекса фиксации Райта (F_{ST}) 2.1 % ($P < 0.05$). Оба использованных метода, главных координат (GenAlEx v.6.503) и кластерный анализ (STRUCTURE 2.3), указывают на несколько бо̀льшую изменчивость выборки быков по сравнению с коровами. При этом на первые главные оси приходится 35 % изменчивости, распределенной примерно поровну. Кластерный анализ показал наличие двух одинаковых кластеров в обеих выборках при значительной выраженности (70 %) у быков кластера, с большими различиями особей друг от друга в противоположность коровам, у которых его доля не более 30 %. Анализ ассоциаций исследованных SNP быков с продуктивными признаками дочерей обнаружил достоверные зависимости ($P < 0.05$) содержания (%) и выхода (кг) жира, удоя (кг) от генотипов гена *SCD* и содержания (%) жира от гена *CSN*3.

Выводы: Определены частотные характеристики исследованных SNP генов кандидатов в выборках обоих полов костромской породы крупного рогатого скота. Поскольку в селекции эффективно использовать аллели с частотой более 5 %, отметим, что аллели большинства маркеров преодолели этот порог, кроме *T*-аллеля гена *Lep* (Y7F) у коров (0.03), что не исключает возможности применения этих маркеров при отборе на продуктивные признаки, поскольку у быков частота данного аллеля составляет 0.07. У костромской породы установленные ассоциативные связи маркеров генов *SCD* (A293V) и *CSN*3 (Ile136Thr и Ala148Asp) с количественными признаками молочной продуктивности дочерей исследованных быков можно учитывать наряду с методами традиционной селекции как при использовании спермы изученных быков для осеменения с коровами неизвестного генотипа, так и для формирования пар животных с известными генотипами при скрещивании для получения потомства с более выраженными признаками.

Финансирование: Исследование поддержано в рамках раздела государственного задания 2024 года ИБР РАН № 0088-2024-0011 и ИЭА РАН № 1023032600025-8-6.1.1 # 5.

Polymorphism of a number of candidate genes for economically important quantitative traits in the Kostroma cattle breed and its association with productive traits

Lazebnaya I.V.[1]*, Lazebny O.E.[2]

[1] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

[2] *Koltzov Institute of Developmental Biology, RAS, Moscow, Russia*

* *Lazebnaya@mail.ru*

Key words: cattle; candidate genes; productivity traits

Motivation and Aim: Selection of livestock animals based on the use of association of alleles of certain genes with economically significant quantitative and qualitative traits (marker-assisted selection, MAS) is one of the effective methods to achieve high results in agricultural production. The search and investigation of gene spectra related to potential physiological and positional DNA markers have been conducted since the end of the last century [1]. Significant results have been achieved. Gene spectra have been identified, the products of which are involved in the formation of certain quantitative traits of dairy and beef productivity in cattle [2]. Despite significant results achieved in the search for new and obtaining data for already known SNP markers, the universal nature of the established associations and their potential use in the selection of all breeds, regardless of the direction of productivity, origin, and achievements of previous selection, remain somewhat unclear. In addition, the frequency characteristics of the distribution of the most valuable alleles in within-breed samples of cattle are important and should be evaluated, as well as the genotyping of sire bulls, prior to conducting the corresponding selection. The aim of this study was to investigate the genetic structure in samples of the Kostroma cattle breed of both sexes using SNP markers of candidate genes (*GH*, *GHR*, *DGAT*1, *PRL*, *SCD*, *RORC*, *Lep*, *LepR*, *CSN*3) and to identify their associations with milk productivity traits of the offspring of the studied bulls.

Methods and Algorithms: Genotyping was conducted on bulls (N = 75) and cows (N = 120) of the Kostroma breed using SNP markers of candidate genes: growth hormone (*GH*, L127V) and its receptor (*GHR*, S555G), diacylglycerol acyltransferase (*DGAT*1, K232A), prolactin (*PRL*, g.35108342 A>G) and kappa-casein (*CSN*3, Ile136Thr and Ala148Asp), stearoyl-CoA desaturase (*SCD*, A293V) retinoic acid C-receptor (*RORC*, g.3984A>G), stearoyl-CoA desaturase (*SCD*, A293V), leptin (*Lep*, Y7F) and its receptor *LepR*, T945M) as described [3–6]. Population genetic analysis was performed using GenAlEx v.6.5 and STRUCTURE 2.3.4 software. Associative analysis of milk productivity traits (milk yield (kg), fat content (%), and fat yield (kg)) of daughters, totaling over 950, was conducted in relation to the genotypes of their fathers, the sire bulls, in STATISTICA 10 software.

Results: In the analyzed samples, no significant deviation was detected in the distribution of observed genotype frequencies from their distribution according to Hardy–Weinberg equilibrium. Differentiation of the samples from each other was identified for the *CSN*3, *DGAT*I and *LepR* gene markers based on the provided probability values for G_{ST} values (Table 1) and considering the Benjamini–Hochberg correction for multiple comparisons [7]. The level of differentiation based on a complex of markers using Wright's fixation index (F_{ST}) was 2.1 % ($P < 0.05$). Both methods used, principal coordinate analysis (GenAlEx v.6.503) and cluster analysis (STRUCTURE 2.3),

indicate slightly greater variability in the bull sample compared to the cows. The first principal axes account for 35 % of the variability, distributed approximately equally. Cluster analysis revealed the presence of two identical clusters in both samples, with a significant predominance (70 %) in the bull cluster, exhibiting significant differences among individuals, as opposed to cows, where its share is no more than 30 %. Analysis of associations of the studied SNPs in bulls with productive traits of daughters revealed significant dependencies ($P < 0.05$) of fat content (%) and fat yield (kg), milk yield (kg) on genotypes of *SCD* gene and fat content (%) on *CSN*3 gene.

Table 1. Values of the G_{ST} coefficient and their probability

Locus	*CSN3*	*PRL*	*GH*	*DGATI*	*SCD*	*RORC(A)*	*GHR*	*Lep*	*LepR*	Tot
Gst	0.016	–0.002	–0.002	0.044	0.007	–0.002	–0.002	0.006	0.034	0.011
P	0.006	0.748	0.627	0.001	0.039	0.957	0.663	0.067	0.001	0.001

Conclusion: Frequency characteristics of the studied SNPs of candidate genes in samples of both sexes of the Kostroma cattle breed were determined. Since it is effective in selection to use alleles with frequency of more than 5 %, we note that alleles of the majority of markers have exceeded this threshold, except for the *T*-allele of *Lep* (Y7F) gene in cows (0.03), which does not exclude the possibility of using these markers in selection for productive traits, as the frequency of this allele in bulls is 0.07. In the Kostroma breed, the established associative relationships of markers of *SCD* (A293V) and *CSN*3 (Ile136Thr and Ala148Asp) genes with quantitative traits of milk productivity of the daughters of the studied bulls can be taken into account along with the methods of traditional breeding, both when using semen of the studied bulls for insemination with cows of unknown genotype and for forming pairs of animals with known genotypes in crossbreeding to obtain offspring with more pronounced traits.

Funding: The study is supported by Government program of basic research in Koltzov Institute of Developmental Biology of the Russian Academy of Sciences No. 0088-2024-0011 and Miklukho-Maklai Institute of Ethnology and Anthropology of the Russian Academy of Sciences No. 1023032600025-8-6.1.1 # 5 in 2024.

Список литературы/References

1. Dekkers J.C. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J Anim Sci*. 2004;82 E-Suppl:E313-328. doi 10.2527/2004.8213_supplE313x
2. Yudin N.S., Voevoda M.I. Molecular genetic markers of economically important traits in dairy cattle. *Russ J Genet*. 2015;51:506-517. doi 10.1134/S1022795415050087
3. Avilés C., Polvillo O., Peña F., Juárez M., Martínez A.L., Molina A. Associations between DGAT1, FABP4, LEP, RORC, and SCD1 gene polymorphisms and fat deposition in Spanish commercial beef. *J Anim Sci*. 2013;91(10):4571-4577. doi 10.2527/jas.2013-6402
4. Lazebnaya I.V., Perchun A.V., Lhasaranov B.B., Lazebny O.E., Stolpovskiy Y.A. Analysis of GH1, GHR and PRL gene polymorphisms for estimation of the genetic diversity of Buryat and Altai cattle breeds. *Vavilovskii Zhurnal Genetiki i Selektsii=Vavilov Journal of Genetics* 2018;22:734-741. doi 10.18699/VJ18.417
5. Gorlov I.F., Fedunin A.A., Randelin D.A., Sulimova G.E. Polymorphisms of *bGH*, *RORC*, and *DGAT1* genes in Russian beef cattle breeds. *Russ J Genet*. 2014;50:1302-1307. doi 10.1134/S1022795414120035
6. Сулимова Г.Е., Лазебная И.В., Перчун А.В. и др. Уникальность костромской породы крупного рогатого скота с позиции молекулярной генетики *Достижения науки и техники АПК*. 2011;9:52-54 [Sulimova G.E., Lazebnaja I.V., Perchun A.V., Voronkova V.N., Ruzina M.N., Badin G.A Uniqueness of Kostroma breed of cattle from a position of molecular genetics. *Dostizheniya nauki i tekhniki APK*. 2011;9:52-54 (in Russia)]
7. Benjamini Y., Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Stat Methodol*. 1990;V(557):289-300. doi 10.1111/j.2517-6161.1995.tb02031.x

doi 10.18699/bgrs2024-5.1-08

Филогеография шерстистого мамонта (*Mammuthus primigenius*) на географически изолированных территориях Восточной Сибири в позднем плейстоцене

Модина С.[1]*, Куслий М.[1], Маликов Д.[2], Павлова Н.[3], Павлов И.[3], Протопопов А.[3], Молодцева А.[1, 4]

[1] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[2] *Институт геологии и минералогии им. В.С. Соболева СО РАН, Новосибирск, Россия*

[3] *Академия наук Республики Саха (Якутия), Якутск, Россия*

[4] *Институт археологии и этнографии СО РАН, Новосибирск, Россия*

* *s.modina@g.nsu.ru*

Ключевые слова: древняя ДНК; шерстистый мамонт; филогеография; митохондриальный геном

Мотивация и цель: К настоящему времени собрана достаточная научная база для определения филогеографии одного из самых многочисленных в летописи окаменелостей представителей мамонтовой фауны – вида шерстистый мамонт (*Mammuthus primigenius*), начиная с анализа частей митохондриального генома и заканчивая изучением полных ядерных геномов [1–3]. Изменение климата от позднего ледникового периода до голоцена привело к уменьшению открытых пространств в Евразии, что привело к сокращению области обитания мамонтов и других степных животных. Этот процесс включал сложные изменения климата и растительности в пространстве и времени, выживание видов в укрытиях, локальное исчезновение и временное расширение областей обитания (Lister, Stuart, 2008). Одним из эффективных подходов для детальной реконструкции этих процессов является исследование локальных серий образцов митохондриальной ДНК мамонтов, относящихся к различным хронологическим периодам. Несмотря на большую степень изученности, выборка образцов из Сибири и Дальнего Востока была представлена в основном северными и восточными регионами. Образцы мамонтовой фауны таких географически изолированных территорий, как Минусинская котловина и остров Котельный, не были исследованы на молекулярно-генетическом уровне, хотя полученные данные могут выявить не обнаруженное ранее генетическое разнообразие вида, уникальные популяции, которые эволюционировали здесь независимо от всех остальных.

Методы и алгоритмы: В рамках данной работы были получены митогеномные библиотеки двенадцати шерстистых мамонтов Минусинской котловины юга Восточной Сибири и десяти шерстистых мамонтов острова Котельный севера Восточной Сибири. Для данных библиотек было проведено двухраундное обогащение геномных библиотек с использованием гибридизации с биотинилированными фрагментами современной митохондриальной ДНК *Elephas maximus*, на основе ранее опубликованного метода [4] с модификациями, описанными в статье Воробьевой и соавторов [5]. Также было проведено высокопроизводительное секвенирование для исследованных образцов. По результатам секвенирования было выполнено множественное выравнивание последовательностей митогеномов позднеплейстоценовых шерстистых мамонтов изолированных территорий

Восточной Сибири и ранее опубликованных последовательностей митогеномов шерстистых мамонтов, проведена филогенетическая реконструкция и построено филогенетическое древо байесовским методом. Для образцов шерстистых мамонтов с Минусинской котловины было построено филогенетическое древо при помощи программной платформы BEAST с внутренней калибровкой времени расхождения ветвей на основе радиоуглеродных датировок образцов для всех митохондриальных последовательностей шерстистых мамонтов, ширина покрытия геномов которых превышала 68 %.

Результаты: Средняя глубина покрытия митогеномов шерстистых мамонтов из Минусинской котловины варьирует от 0.5 до 15.5 раза, а ширина покрытия составляет от 38 до 99.5 % от длины референсного митогенома. Для митогеномов шерстистых мамонтов с острова Котельный ширина покрытия составляет от 16 до 97.1 % от длины референсного митогенома. Исходя из полученных значений частоты дезаминирования оснований и среднего размера фрагментов ДНК, можно заключить, что костные образцы мамонтов имеют высокую степень сохранности ДНК, что связано с относительно хорошими для сохранности ДНК условиями окружающей среды.

Выявлено, что последовательности мтДНК шерстистых мамонтов Минусинской котловины и острова Котельный не формируют отдельную кладу на дереве, а рассредоточены в разных кластерах клады I, что отличает их от некоторых других исследованных локальных групп мамонтов, таких как мамонты с острова Врангеля, находившихся в стадии сниженного генетического разнообразия. Филогеографические реконструкции на основе последовательностей митохондриальной ДНК шерстистых мамонтов из Минусинской котловины подтвердили время расхождения I и II митохондриальных клад мамонтов, как 2–1 млн лет, выявили генетическую близость митохондриальных линий позднеплейстоценовых мамонтов Минусинской котловины и других, в том числе сопредельных, регионов Восточной Сибири и их дивергенцию во временном промежутке от 150 до 100 тыс. лет назад, что свидетельствует об активных миграциях шерстистых мамонтов на обширных территориях Восточной Сибири (острова Врангеля, Центрального, Западного и Восточного регионов) в конце среднего плейстоцена–начале позднего плейстоцена.

Выводы: Данное исследование позволяет оценить митохондриальное генетическое разнообразие шерстистых мамонтов на территории Минусинской котловины и острова Котельный. Расположение митотипов в разных кладах внутри клады I может указывать на достаточно высокое разнообразие генофонда исследованных шерстистых мамонтов. По проведенным филогеографическим реконструкциям можно говорить об активных миграциях шерстистых мамонтов на обширных территориях Восточной Сибири в конце среднего плейстоцена–начале позднего плейстоцена.

Финансирование: Исследование поддержано грантом РНФ (№ 23-74-10060).

Phylogeography of the woolly mammoth (*Mammuthus primigenius*) in geographically isolated areas of Eastern Siberia in the Late Pleistocene

Modina S.[1]*, Kusliy M.[1], Malikov D.[2], Pavlova N.[3], Pavlov I.[3], Protopopov A.[3], Molodtseva A.[1, 4]

[1] *Institute of Molecular and Cell Biology, SB RAS, Novosibirsk, Russia*

[2] *V.S. Sobolev Institute of Geology and Mineralogy, SB RAS, Novosibirsk, Russia*

[3] *Academy of Sciences of the Republic of Sakha (Yakutia), Yakutsk, Russia*

[4] *Institute of Archaeology and Ethnography, SB RAS, Novosibirsk Russia*

* *s.modina@g.nsu.ru*

Key words: ancient DNA; woolly mammoth; phylogeography; mitochondrial genome

Motivation and Aim: To date, a sufficient scientific basis has been assembled to determine the phylogeography one of the most numerous representatives of the mammoth fauna, the woolly mammoth (*Mammuthus primigenius*) species, in the fossil record, from the analysis of parts of the mitochondrial genome to the study of complete nuclear genomes [1–3]. Climate change from the LGP to the Holocene led to a decrease in open spaces in Eurasia, resulting in a shrinking habitat for mammoths and other steppe animals. This process involved complex changes in climate and vegetation in space and time, survival of species in shelter, localized extinction, and temporary expansion of habitat areas (Lister, Stuart, 2008). One effective approach for detailed reconstruction of these processes is the study of localized series of mammoth mitochondrial DNA samples from different chronological periods. Despite the large degree of study, the sampling of samples from Siberia and the Far East was represented mainly by the northern and Eastern regions. Samples of the mammoth fauna of such geographically isolated areas as the Minusinsk Depression and Kotelny Island have not been studied at the molecular genetic level, although the data obtained may reveal a previously undiscovered genetic diversity of the species, unique populations that evolved here independently of all others. *Methods and Algorithms*: Mitogenomic libraries for twelve woolly mammoths from the Minusinsk Depression in southern Eastern Siberia and ten woolly mammoths from Kotelny Island in northern Eastern Siberia were generated in this study. For these libraries, based on a previously published method [4] with modifications described in Vorobieva et al. [5], two-round enrichment of genomic libraries was performed by hybridisation with biotinylated fragments of modern mitochondrial DNA from *Elephas maximus*. The investigated samples were also subjected to high-throughput sequencing. Based on the sequencing results, multiple alignment of mitogenome sequences of Late Pleistocene woolly mammoths from isolated areas of Eastern Siberia and previously published mitogenome sequences of woolly mammoths was performed, phylogenetic reconstruction was performed, and a phylogenetic tree was constructed using the Bayesian method. For woolly mammoth samples from the Minusinsk Depression, a phylogenetic tree was constructed using the BEAST software platform with internal calibration of branch divergence time based on radiocarbon dating of samples. The tree was constructed for all woolly mammoth mitochondrial sequences with genome coverage greater than 68 %.

Results: The average depth of coverage of the mitogenomes of woolly mammoths from the Minusinsk Depression varies from 0.5 to 15.5 times, and the width of coverage ranges from 38 to 99.5 % of the reference mitogenome length. The coverage width ranges from 16 to 97.1 % of the reference mitogenome length for woolly mammoths from Kotelny Island. Based on the obtained values of base deamination frequency and average size of DNA fragments, it can be concluded that mammoth bone samples have a high degree of DNA preservation. This is associated with relatively good environmental conditions for DNA preservation.
The mtDNA sequences of woolly mammoths from the Minusinsk Depression and Kotelny Island, unlike some other local groups of mammoths studied, such as mammoths from Wrangel Island, which were in a stage of reduced genetic diversity, were found not to form a separate clade on the tree but to be distributed in different clusters of clade I. Phylogeographic reconstructions based on mitochondrial DNA sequences of woolly mammoths from the Minusinsk Depression confirmed the time of divergence of the I and II mammoth mitochondrial clades as 2–1 million years ago, revealed the genetic proximity of mitochondrial lineages of Late Pleistocene mammoths from the Minusinsk Depression and other, including adjacent, regions of Eastern Siberia, and their divergence in the time interval from 150 to 100 thousand years ago, which points to active migrations of woolly mammoths in the vast territories of Eastern Siberia (the Wrangel Islands, the central, western and Eastern regions) in the late Middle Pleistocene – early Late Pleistocene.
Conclusion: This study allows an assessment of the mitochondrial genetic diversity of woolly mammoths in the Minusinsk Depression and Kotelny Island. The location of mitotypes in different clades within clade I may indicate a rather high diversity of the gene pool of the studied woolly mammoths. According to phytogeographic reconstructions, woolly mammoths were actively migrating across the vast territories of Eastern Siberia at the end of the Middle Pleistocene – beginning of the Late Pleistocene.
Funding: The study is supported by Russian Science Foundation (No. 23-74-10060).

Список литературы/References

1. Gilbert et al. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science.* 2007;317(5846):1927-1930
2. Palkopoulou E., Dalén L., Lister A.M. et al. Holarctic genetic structure and range dynamics in the woolly mammoth. *Proc Biol Sci.* 2013;280(1770):20131910. doi 10.1098/rspb.2013.1910
3. Van der Valk T. et al. Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* 2021;591(7849):265-269. doi 10.1038/s41586-021-03224-9
4. Maricic T., Whitten M., Pääbo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*. 2010;5(11):e14004. doi 10.1371/journal.pone.0014004
5. Vorobieva N.V. et al. High genetic diversity of ancient horses from the Ukok Plateau. *PLoS One.* 2020;15(11):e0241997. doi 10.1371/journal.pone.0241997

doi 10.18699/bgrs2024-5.1-09

Вовлеченность микроРНК в процесс доместикации у представителей рода *Bos*

Скобель О.И.*, Глазко В.И., Косовский Г.Ю.
Научно-исследовательский институт пушного звероводства и кролиководства им. В.А. Афанасьева, Московская область, Раменский район, Россия
** skobelolga@gmail.com*

Ключевые слова: крупный рогатый скот; бизон; ретротранспозоны; микроРНК; доместикация; mir-30; RTE-BovB; ERV

Мотивация и цель: Для понимания эволюции генетических механизмов сложных фенотипических признаков необходимо выявление ключевых элементов регуляции их изменчивости путем сравнения близкородственных видов [1]. Одним из компонентов регуляторных сетей профилей генной экспрессии являются микроРНК, вовлекаемые в процессы одомашнивания и последующие этапы искусственного отбора [2]. Основным источником возникновения новых микроРНК, а также их распространения по геному являются ретротранспозоны (ТЕ) [3]. В рамках наших предыдущих исследований по анализу распределения и позиционирования на 1-й хромосоме крупного рогатого скота (КРС) выявлены 30 конструкций вида RTE-BovB/ERVK/RTE-BovB, локализованных в интронных областях 12 структурных генов. Указанные гены образуют тесно сцепленный блок, сохраняющийся в эволюции млекопитающих [4]. Нуклеотидная консервативность этого продукта рекомбинации трех ТЕ позволяет предполагать наличие в нем функциональных элементов, обеспечивающих его сохранение [5]. Для выяснения степени вовлеченности в процесс доместикации продуктов рекомбинации ТЕ выполнено их прямое сравнение *in silico* у крупного рогатого скота (*Bos taurus*) и бизона (*Bison bison*), а также их возможной продукции как источника микроРНК в синтенном блоке генов.
Методы и алгоритмы: Нуклеотидные последовательности 30 конструкций вида RTE-BovB/ERVK/RTE-BovB у КРС получены нами ранее в предыдущих исследованиях [4, 5]. Для целей данной работы использована сборка генома бизона – Bison_UMD1.0 [6]. Анализ сходства нуклеотидных последовательностей проводили с использованием программ Clustal Omega и Kalign при множественном выравнивании с настройками по умолчанию [7, 8]. Проверку наличия микроРНК проводили с использованием возможностей базы данных микроРНК, The microRNA database [9].
Результаты: Установлено, что вышеуказанные продукты рекомбинации с высокой степенью сходства (не ниже 81.5 %) представлены в соответствующем блоке из 12 генов у бизона, но имеют пониженную частоту встречаемости по сравнению с КРС. Они также являются источниками 51 микроРНК у бизона и 129 – у крупного рогатого скота. Выяснилось, что 50 микроРНК являются общими как для КРС, так и для бизона. В то же время одна микроРНК, cli-miR-1416-3p, отсутствовала у крупного рогатого скота, но присутствовала у бизона. 79 микроРНК отсутствовали у бизона, но присутствовали у крупного рогатого скота, в том числе 78 микроРНК, относящихся к семейству miR-30, включая две

специфические для крупного рогатого скота микроРНК: bta-miR-30a-5p и bta-miR-30e-5p. Накоплены литературные данные о том, что представители этого семейства участвуют в увеличении жирности молока, играют ключевую роль в развитии мышечной ткани у крупного рогатого скота и вовлечены в реакцию на стресс и развитие иммунного ответа [10–12].
Выводы: Таким образом, можно предположить, что продукты трехчленной рекомбинации между LINE BovB и ERV активно вовлекаются в доместикацию, являясь источниками микроРНК, оказывающими значительное влияние на важные сельскохозяйственные признаки крупного рогатого скота. Из этого следует, что при доместикации регуляторные сети могут существенно меняться не только за счет возникновения новых сайтов посадки микроРНК [2], но и за счет формирования новых вариантов микроРНК.
Финансирование: Исследование поддержано Министерством науки и высшего образования Российской Федерации (№ 075-00503-24-01).

MicroRNAs' involvement in the domestication process of the genus *Bos* members

Skobel O.I.*, Glazko V.I., Kosovsky G.Yu.
Afanas'ev Research Institute of Fur-Bearing Animal Breeding and Rabbit Breeding, Moscow region, Ramensky district, Russia
** skobelolga@gmail.com*

Key words: cattle; bison; retrotransposons; microRNA; mir-30; domestication; RTE-BovB; ERV

Motivation and Aim: Understanding the evolution of genetic mechanisms of complex phenotypic traits becomes possible by identifying key elements of their variability regulation by comparing closely related species [1]. MicroRNAs are one of the key elements of gene expression regulatory networks and are also involved in the processes of domestication and subsequent stages of artificial selection [2]. Retrotransposons (TEs) are the main source of the emergence of new microRNAs as well as their genome-wide distribution [3]. Previously, we identified 30 RTE-BovB/ERVK/RTE-BovB constructs localized in the intronic regions of 12 structural genes on the bovine chromosome 1. The above genes form a closely linked block that is conserved in mammalian evolution [4]. The nucleotide conservativity of this trinominal recombination product suggests that it contains functional elements that ensure its conservation [5]. To clarify the degree of involvement of TE recombination products in the process of domestication, their direct *in silico* comparison in cattle (*Bos taurus*) and bison (*Bison bison*) genomes was performed in the syntenic gene block. There, the possible existence of microRNA was also examined.
Methods and Algorithms: Compared nucleotide sequences of 30 bovine RTE-BovB/ERVK/RTE-BovB constructs were obtained from our previous studies [4, 5]. The bison gene data were based on the Bison_UMD1.0 assembly [6]. We used Clustal Omega [7] and Kalign [8] for multiple alignments with default settings. The presence of microRNAs was checked using the microRNA database [9].
Results: The above recombination products with a high percent of identity (at least 81.5 %) are found in the corresponding block of 12 genes in bison but with a lower frequency compared to cattle. They also provide 51 microRNAs in bison and

129 microRNAs in cattle. 50 microRNAs are common to both cattle and bison. At the same time, one microRNA, cli-miR-1416-3p, was absent in cattle but present in bison. There were 79 microRNAs that were present in cattle but not in bison. Of these, 78 microRNAs were from the miR-30 family, including two bovine-specific microRNAs: bta-miR-30a-5p and bta-miR-30e-5p. Representatives of this family are known to be involved in increasing milk fat content [10], play a key role in the development of muscle tissue in cattle [11], and are involved in stress and immune response development [12].

Conclusion: Given that the trinominal recombination products between LINE BovB and ERV are sources of microRNAs that have significant effects on important agricultural traits in cattle, it can be assumed that these products are actively involved in domestication. This implies that regulatory networks can change significantly during domestication, not only through the emergence of new microRNA binding sites [2], but also through the formation of new microRNA variants.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (No. 075-00503-24-01).

Список литературы/References

1. Chen S., Liu S., Shi S. et al. Comparative epigenomics reveals the impact of ruminant-specific regulatory elements on complex traits. BMC Biol. 2022;20:273. doi 10.1186/s12915-022-01459-0
2. Braud M., Magee D.A., Park S.D.E. et al. Genome-wide microRNA binding site variation between extinct wild aurochs and modern cattle identifies candidate microRNA-regulated domestication genes. Front Genet. 2017;8:3. doi 10.3389/fgene.2017.00003
3. Roberts J.T., Cardin S.E., Borchert G.M. Burgeoning evidence indicates that microRNAs were initially formed from transposable element sequences. Mob Genet Elements. 2014;4:e29255. doi 10.4161/mge.29255
4. Глазко В.И., Скобель О.И., Косовский Г.Ю., Глазко Т.Т. Доменная организация мобильных генетических элементов в 1-й хромосоме крупного рогатого скота. Сельскохозяйственная биология 2017;52(4):658-668.
   [Glazko V.I., Skobel O.I., Kosovsky G.Yu., Glazko T.T. Domain distribution of mobile genetic elements in the bovine genome. Sel'skokhozyaistvennaya biologiya = Agricultural Biology. 2017;52(4):658-668 (in Russian)]
5. Скобель О.И., Глазко В.И., Косовский Г.Ю., Глазко Т.Т. Рекомбинации между мобильными генетическими элементами как источник микроРНК. Известия ТСХА. 2017;(4):70-98. doi 10.26897/0021-342X-2017-4-70-98
   [Skobel O.I., Glazko V.I., Kosovsky G.Yu., Glazko T.T. Recombinations among mobile genetic elements as a source of microRNA. Izv Timirâzevsk s-h akad. 2017;(4):70-98. doi 10.26897/0021-342X-2017-4-70-98 (in Russian)]
6. Achilli A., Olivieri A., Pellecchia M. et al. Mitochondrial genomes of extinct aurochs survive in domestic cattle. Curr Biol. 2008;18(4):R157-R158. doi 10.1016/j.cub.2008.01.019 1
7. Sievers F., Wilm A., Dineen D. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7(1):539. doi 10.1038/msb.2011.75
8. Lassmann T., Frings O., Sonnhammer E.L.L. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. Nucleic Acids Res. 2009;37(3):858-865. doi 10.1093/nar/gkn1006
9. Kozomara A., Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42(Database issue):D68-D73. doi 10.1093/nar/gkt1181
10. Chen Z., Qiu H., Ma L. et al. miR-30e-5p and miR-15a synergistically regulate fatty acid metabolism in goat mammary epithelial cells via LRP6 and YAP1. IJMS. 2016;17(11):1909. doi 10.3390/ijms17111909
11. Eisenberg I., Eran A. et al. Distinctive patterns of microRNA expression in primary muscular disorders. Proc Natl Acad Sci USA. 2007;104(43):17016-17021. doi 10.1073/pnas.0708115104
12. Zheng Y., Chen K., Zheng X., Li H., Wang G. Identification and bioinformatics analysis of microRNAs associated with stress and immune response in serum of heat-stressed and normal Holstein cows. Cell Stress Chaperones. 2014;19(6):973-981. doi 10.1007/s12192-014-0521-8

doi 10.18699/bgrs2024-5.1-10

Полногеномные данные подтверждают присутствие кроссинговера у гибридов соболя (*M. zibellina*) и лесной куницы (*M. martes*)

Томаровский А.[1, 2]*, Тотиков А.[1, 2], Беклемишева В.[1], Перельман П.[1], Сердюкова Н.[1], Бульонкова Т.[3], Сидоров М.[4], Мамаев Н.[4], Охлопков И.[4], Мухачева А.[5], Коняева К.[6], Абрамов А.[7], Графодатский А.[1], Кливер С.[8]

[1] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Институт систем информатики им. А.П. Ершова СО РАН, Новосибирск, Россия*

[4] *Институт биологических проблем криолитозоны СО РАН, Якутск, Россия*

[5] *Сихотэ-Алинский государственный природный биосферный заповедник им. К.Г. Абрамова, п. Терней, Россия*

[6] *Зоопарк «Лесная сказка», Барнаул, Россия*

[7] *Зоологический институт РАН, Санкт-Петербург, Россия*

[8] *Независимый исследователь, Копенгаген, Дания*

* *andrey.tomarovsky@gmail.com*

Ключевые слова: гибридизация; анализ генетического смешения

Мотивация и цель: Соболь (*M. zibellina*) и лесная куница (*M. martes*) являются близкородственными представителями рода *Martes* (сем. Mustelidae). В зоне перекрывания ареалов этих пушных видов происходит образование гибридов – кидасов, которые характеризуются промежуточными фенотипическими признаками родительских видов [1]. Внутрипопуляционный и межвидовой анализ генетической структуры этих симпатрических видов ранее проводился с использованием подходов низкого разрешения, основанных на анализе митохондриальных и отдельных ядерных маркеров [2, 3]. Используя собственные полногеномные данные 7 образцов *M. martes*, 13 образцов *M. zibellina* и 12 образцов их предполагаемых гибридов, а также полногеномные данные 1 опубликованного образца китайского подвида соболя, мы выполнили ADMIXTURE-анализ исследуемых особей.

Методы и алгоритмы: Поиск, маскирование и фильтрация SNP аутосом и псевдоаутосомного района выполнены при помощи BCFTOOLS, BEDTOOLS, BEDOPS и PLINK [4–7]. В качестве референса использована общедоступная геномная сборка соболя хромосомного уровня [8]. Анализ главных компонент (PCA) проведен с помощью PLINK. Визуализация полученного результата и вычисление объясненной дисперсии каждой компоненты выполнены средствами языка программирования R. Анализ генетического смешения образцов выполнен с использованием ADMIXTURE [8] для 1–6 предполагаемых генетических кластеров исходных популяций (K). Анализ проводился в трехкратной повторности с вычислением ошибок перекрестной валидации и значений среднего попарного сходства. Локальный ADMIXTURE анализ выполнен путем отдельных запусков ADMIXTURE (K = 2) с использованием неперекрывающихся скользящих окон

размером 1 млн п. н. Визуализация выполнена средствами языка программирования Python.

Результаты: На графике PCA (рис. 1, A) исследуемые образцы сформировали две группы относительно первой главной компоненты. Левая группа сформирована лесными куницами. В свою очередь правая группа, особи из которой больше рассредоточены по вертикали, состоит из соболей. Образцы T87 и T84, а также в меньшей степени T18 расположились практически между двумя группами, что является признаком наиболее выраженного гибридного статуса этих особей.

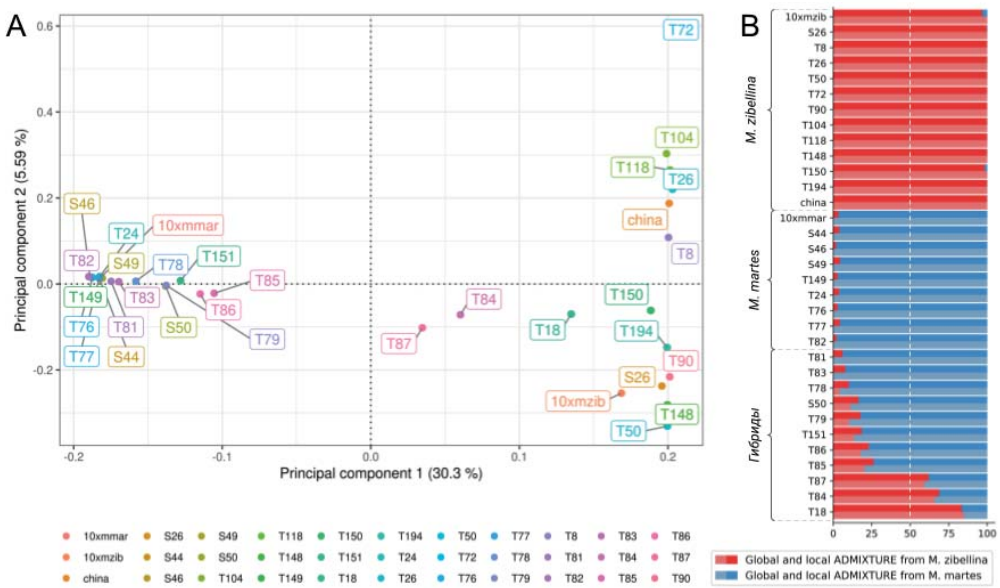


Рис. 1. A – график PCA исследуемых образцов соболя, лесной куницы и их предполагаемых гибридов; B – результаты глобального и локального ADMIXTURE-анализа (K = 2)

Результаты глобального и локального ADMIXTURE-анализа (K = 2) визуализированы на рис. 1, B. Глобальный ADMIXTURE-анализ выявил наличие интрогрессии в 9 образцах, доля которой колеблется в диапазоне от 3.9 % (T78) до 40.8 % (T87). Следов интрогрессии среди остальных образцов обнаружено не было. Локальный ADMIXTURE-анализ позволил дополнительно выявить интрогрессию в образцах T81 и T83, равную 6.17 и 7.86 % соответственно. В общей сложности мы обнаружили следы интрогрессии от 1 до 5 % в 11 образцах (включая референсные образцы родительских видов).

Локальный ADMIXTURE-анализ на уровне каждой хромосомы дополнительно позволил нам установить мозаичность хромосом у гибридных особей. В качестве примера на рис. 2, A показан гибридный образец T18, на 10-й хромосоме которого выявлены как участки от каждого из родительских видов, так и гибридные участки от обоих из них. Мозаичность представленного и других гибридных образцов проявляется в разных пропорциях и может проявляться на разных хромосомных участках, что свидетельствует о том, что кроссинговер в гибридных образцах не подавлен. Дополнительно показаны две подтвержденных лесных куницы из Свердловской (рис. 2, B) и Архангельской (рис. 2, C) областей с интрогрессией со

стороны соболя в 3.4 и 4.3 % соответственно. Подобные следы интрогрессии у типичных особей лесной куницы говорят о расширении западной части ареала соболя, а также, вероятно, могут быть следствием побегов соболей со звероферм.

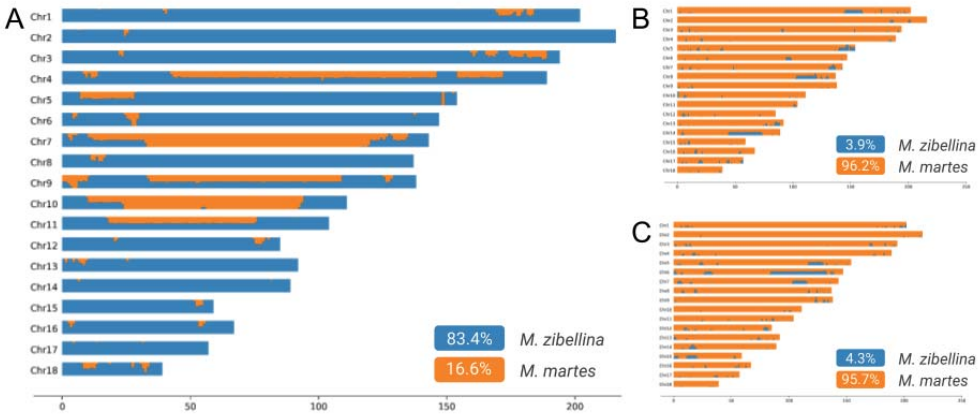


Рис. 2. Результаты локального ADMIXTURE-анализа каждой хромосомы: A – для гибридного образца T18; B и C – для подтвержденных образцов лесной куницы

Выводы: Нам удалось выявить взаимные интрогрессии на уровне каждой хромосомы между соболем и лесной куницей. Активный двунаправленный поток генов наблюдается в геномах особей как из зоны симпатрии, так и в удаленных от этой области частях ареалов родительских видов. Среди образцов лесных куниц не обнаружено ни одной особи без интрогрессии от соболя, что является свидетельством расширения западной части ареала соболя. Из рассмотренных 12 образцов предполагаемых кидасов подтвержденными оказались только 8, тогда как оставшиеся 4 являются лесными куницами с интрогрессией со стороны соболя менее условного порога в 5 %. Полученные результаты подтверждают, что кроссинговер у гибридных особей не подавлен.

Благодарности: Выражаем благодарность Александру Мигуру, Валерию Зелепухину, Виктору Хазагаеву, Валерию Середкину, Валентину Слободенюку, Сергею Крехнову, Ирине Шуваловой, Светлане Зубакиной, Алексею Лахтину, Сергею Цыпленкову, Сергею Писареву, Павлу Резниченко и Валерию Жданюк за помощь в сборе образцов.

Финансирование: Исследование поддержано грантом РНФ № 19-14-00034-П.

Whole-genome data confirm the presence of crossover in hybrids of sable (*M. zibellina*) and pine marten (*M. martes*)

Tomarovsky A.[1, 2]*, Totikov A.[1, 2], Beklemisheva V.[1], Perelman P.[1], Serdyukova N.[1], Bulyonkova T.[3], Sidorov M.[4], Mamaev N.[4], Okhlopkov I.[4], Mukhacheva A.[5], Koniaeva K.[6], Abramov A.[7], Graphodatsky A.[1], Kliver S.[8]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *A.P. Ershov Institute of Informatics Systems, SB RAS, Novosibirsk, Russia*

[4] *Institute for Biological Problems of Cryolithozone, SB RAS, Yakutsk, Russia*

[5] *Sikhote-Alin State Nature Biosphere Reserve named after K.G. Abramov, Terney, Russia*
[6] *"Lesnaya skazka" Zoo, Barnaul, Russia*
[7] *Zoological Institute RAS, St. Petersburg, Russia*
[8] *Independent scientist, Copenhagen, Denmark*
* *andrey.tomarovsky@gmail.com*

Key words: hybridization; genetic admixture analysis

Motivation and Aim: Sable (*M. zibellina*) and pine marten (*M. martes*) are closely related species of the genus *Martes* (family Mustelidae). In the area of overlapping ranges of these fur species, hybrids – kidases – are formed, which are characterized by intermediate phenotypic traits of the parental species [1]. Intrapopulation and interspecific analysis of the genetic structure of these sympatric species was previously performed using low-resolution approaches based on the analysis of mitochondrial and individual nuclear markers [2, 3]. Using the obtained whole-genome data of 7 samples of *M. martes*, 13 samples of *M. zibellina* and 12 samples of their putative hybrids, as well as the whole-genome data of 1 published sample of the Chinese subspecies of sable, we performed ADMIXTURE analysis of the studied individuals.
Methods and Algorithms: SNPs of autosomes and pseudoautosomal region were identified, masked, and filtered using BCFTOOLS, BEDTOOLS, BEDOPS, and PLINK [4–7]. The publicly available chromosome-level genome assembly of sable was used as a reference [8]. Principal Component Analysis (PCA) was conducted using PLINK. Visualization of the results and calculation of the explained variance for each component were performed using R. Genetic admixture analysis of the samples was performed using ADMIXTURE [8] for 1–6 assumed genetic clusters of ancestral populations (K). The analysis was performed in 3-fold repetition with calculation of cross validation errors and average pairwise similarity values. Local ADMIXTURE analysis was performed by separate runs of ADMIXTURE (K = 2) using non-overlapping sliding windows of 1 Mbp. Visualization was performed using Python.
Results: In the PCA plot (Fig. 1A), the studied samples formed two groups relative to the 1st principal component. The left group is formed by pine martens, while the right group, whose individuals are more dispersed vertically, consists of sables. Samples T87 and T84, and to a lesser extent T18, are located almost between the two groups, indicating a more pronounced hybrid status of these individuals.
The results of global and local ADMIXTURE analysis (K = 2) are visualized in Fig. 1B. Global ADMIXTURE analysis revealed the presence of introgression in 9 samples, ranging from 3.9 % (T78) to 40.8 % (T87). No traces of introgression were found among the remaining samples. Local ADMIXTURE analysis additionally identified introgression in samples T81 and T83 equal to 6.17 and 7.86 % respectively. In total, we detected traces of introgression from 1 to 5 % in 11 samples (including reference samples of the parental species).

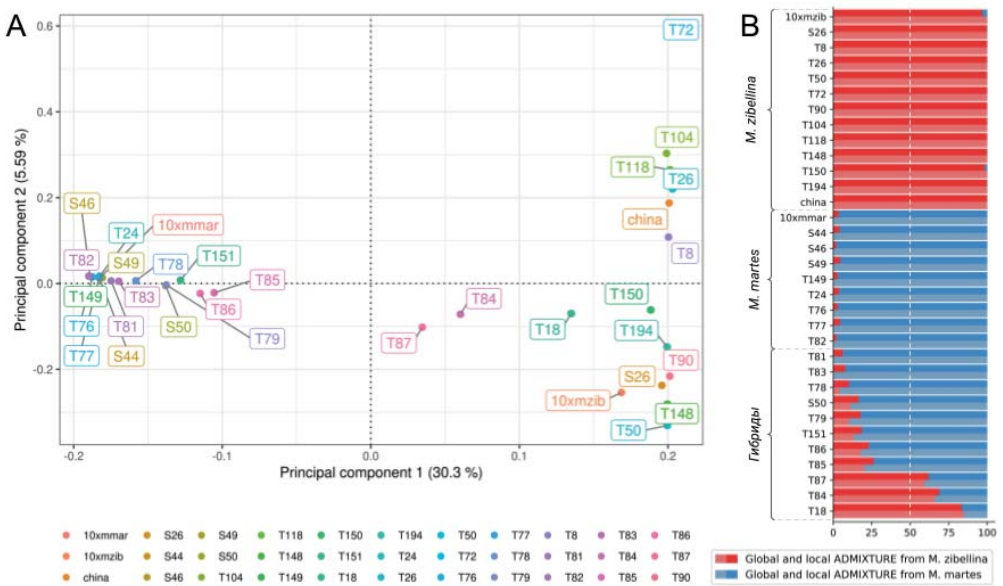


Fig. 1. A, PCA plot of the studied samples of sable (*M. zibellina*), pine marten (*M. martes*) and their putative hybrids; B, results of global and local ADMIXTURE analysis (K = 2)

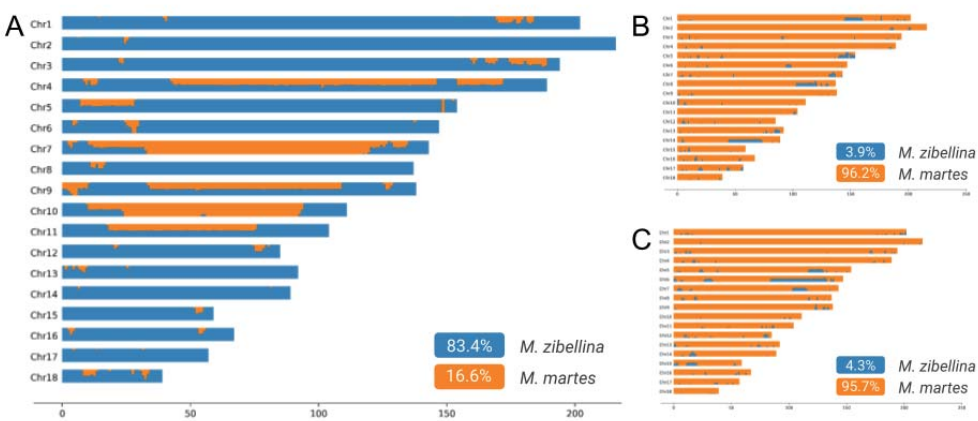


Fig. 2. Results of local ADMIXTURE analysis for each chromosome: A, for hybrid sample T18; B and C, for confirmed pine marten samples

Local ADMIXTURE analysis at the level of each chromosome additionally allowed us to identify chromosome mosaicism in hybrid individuals. As an example, in Fig. 2A, the hybrid sample T18 is shown, where on the 10th chromosome, regions from each of the parental species as well as hybrid regions from both of them were identified. The mosaicism of both the presented and other hybrid samples appears in various proportions and may appear on different chromosomal regions, indicating that crossover is not suppressed in the hybrid samples. Additionally, two confirmed pine marten individuals from Sverdlovsk (Fig. 2B) and Arkhangelsk (Fig. 2C) regions with introgression from sable at 3.4 and 4.3 %, respectively, are shown. Such traces of introgression in typical

pine martens indicate the expansion of the western part of the sable's range and may also be a result of sable escapes from fur farms.

Conclusion: We succeeded in identifying mutual introgressions at the level of each chromosome between sable and pine marten. Active bidirectional gene flow was observed in the genomes of individuals both from the sympatric zone and in remote parts of the ranges of the parental species. Among the pine marten samples, not a single individual without sable introgression was found, indicating the expansion of the western part of the sable's range. Out of the 12 examined suspected hybrids, only 8 were confirmed, while the remaining 4 are pine martens with introgression below the arbitrary threshold of 5 %. The results confirm that crossover is not suppressed in hybrid individuals.

Acknowledgements: We acknowledge Alexander M. Migura, Valerii A. Zelepukhin, Victor D. Hazagaev, Valeriy V. Seredkin, Valentin S. Slobodenuk, Sergei D. Krekhnov, Irina A. Shuvalova, Svetlana V. Zubakina, Aleksey A. Lakhtin, Sergei M. Tsiplenkov, Sergei Pisarev, Pavel Reznichpenko and Valeriy T. Zhdanuk for helping with sample gathering.

Funding: The study is supported by the RSF grant No. 19-14-00034-P.

Список литературы/References

1. Юргенсон П. Кидас – гибрид соболя и куницы. *Труды Печоро-Илычского заповедника*. 1947;5:145-178
   [Yurgenson P. Kidas – a hybrid of sable and marten. *Trudy Pechoro-Ilychskogo zapovednika.* 1947;5:145-178 (in Russian)]
2. Li B. et al. Phylogeography of subspecies of the sable (Martes zibellina L.) based on mitochondrial genomes: implications for evolutionary history. *Mamm Biol.* 2021;101(1):105-120. doi 10.1007/s42991-020-00092-0
3. Пищулина С.Л. Взаимодействие популяций лесной куницы и соболя в зоне симпатрии: генетический аспект. М.: ИПЭЭ РАН, 2013
   [Pishulina S.L. Interaction of pine marten and sable populations in the sympatry zone: genetic aspect. M.: IPEE RAS, 2013 (in Russian)]
4. Danecek P. et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008. doi 10.1093/gigascience/giab008
5. Quinlan A.R., Hall I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi 10.1093/bioinformatics/btq033
6. Neph S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics*. 2012;28(14):1919-1920. doi 10.1093/bioinformatics/bts277
7. Chang C.C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4(1):7. doi 10.1186/s13742-015-0047-8
8. Martes_zibellina. DNA Zoo. https://www.dnazoo.org/assemblies/Martes_zibellina (accessed March 20, 2024) [Electronic resource]
9. Alexander D.H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655-1664. doi 10.1101/gr.094052.109

doi 10.18699/bgrs2024-5.1-11

Последствия 26 поколений отбора норок (*Neogale vison*) по оборонительной реакции на человека

Трапезов О.В.[1, 2]*, Некрасова М.А.[1], Степанова М.А.[1, 3], Баишникова И.В.[4], Илюха В.А.[5], Калинина С.Н.[4], Кижина А.Г.[4], Морозов А.В.[4], Панова Э.В.[4], Балан О.В.[4]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Новосибирский государственный аграрный университет, Новосибирск, Россия*

[4] *Институт биологии КНЦ РАН, Петрозаводск, Россия*

[5] *Институт биологии внутренних вод им. И.Д. Папанина РАН, Борок, Ярославская область, Россия*

* *trapezov@bionet.nsc.ru*

Ключевые слова: селекция; *Neogale vison*; оборонительное поведение; пищеварение; физиологические признаки

Мотивация и цели: В ходе многолетнего эксперимента по селекционному преобразованию оборонительной реакции на человека у американских норок было показано, что отбор открывает новые возможности для наследственной изменчивости в том же направлении, в котором действует сам отбор. Такой отбор способен создавать увеличение темпов и размаха изменчивости не только в направлении своего действия, но и в других направлениях (Беляев, 1968, 1970, 1972, 1974, 1979, 1981, 1983; Беляев, Трут, 1964, 1982, 1983; Trut, 1999, 2007, 2008; Trapezov, 1987, 2012). Цель работы – сравнение темпов и размаха изменчивости в двух линиях норок, селекционируемых на ручное и агрессивное поведение.

Методы и алгоритмы: За 26 поколений отбора в популяции американских норок клеточного разведения генотипа *Standard dark brown* (+/+) были созданы две линии животных с ручным и агрессивным поведением по отношению к человеку. В ходе селекционного преобразования поведения рассчитывались: 1) *селекционный сдвиг* – разность между средними значениями признака у потомства отобранных родителей и в родительском поколении в целом; 2) *селекционный дифференциал* – разность между средними значениями признака у особей, отбираемых в качестве родителей следующего поколения и в целой популяции; 3) *величины реализованной наследуемости* h^2 – насколько величина селекционируемого признака обусловлена наследственностью и насколько влиянием внешней среды; 4) *коэффициенты отбора* – S (%); 5) *популяционные коэффициенты инбридинга по Фальконеру* – F_t ; 6) фактические ответы на отбор (R) на поколение.

Результаты: В ходе эксперимента по селекционному преобразованию оборонительной реакции были выявлены взаимоотношения между направлением отбора и изменчивостью по поведению. Прежде всего, как видно на рис. 1, фенотипическая изменчивость при отборе на доместикационное поведение на два порядка величин выше, чем в группе, селекционируемой на агрессивное поведение. Расчеты среднего показателя фактического ответа на отбор на одно

поколение показывают наибольший эффект отбора по доместицируемым животным. Причем по ним наблюдается половой диморфизм – наибольший эффект отбора зафиксирован по самцам. В экспрессивности ручного поведения (самцы достоверно более ручные, чем самки ($p < 0.001$). Причем величина фенотипической вариансы селекционируемого признака в доместицируемой популяции на два порядка величин выше, чем в популяции, селекционируемой на агрессивность. Средний показатель фактического ответа на отбор на одно поколение по агрессивным животным, как по самцам, так и по самкам, практически одинаковый.

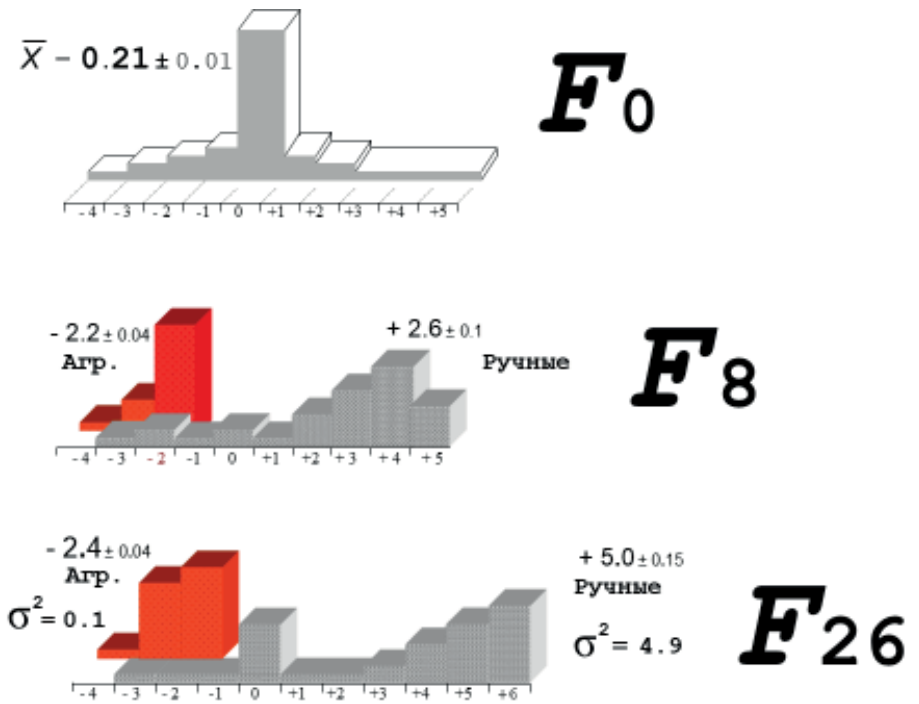


Рис. 1

Реализованная наследуемость, оцененная как отношение сдвига при отборе (R) к селекционному дифференциалу (S), показала, что за ответ на первые два поколения «минус»-отбора ответственны около 60 % аддитивного разнообразия, «плюс»-отбора – около 40 % (37 % для самок и 38 % для самцов). Существенный сдвиг, полученный в течение первых двух поколений отбора на агрессивность, указывает на то, что число основных генов, детерминирующих порог агрессивных реакций, минимально, возможно, не более двух-трех, и они быстро фиксируются на первых этапах отбора.

Отбор по оборонительной реакции на человека приводит к перестройкам регуляции физиологических функций. Животные из агрессивной линии характеризовались более высоким базальным уровнем кортизола в сыворотке крови по сравнению с ручными животными. В то же время ручные особи отличались от агрессивных повышенным содержанием эозинофилов, но сниженным – нейтрофилов, более крупными размерами лимфоцитов и меньшим

ядерно-клеточным соотношением. У ручных норок обнаружен амилолитический профиль активности пищеварительных ферментов и более высокое содержание витаминов *A* и *E* в печени, тогда как агрессивные особи отличались более высокой активностью протеолитических и липолитических ферментов пищеварительного тракта. Можно допустить, что генная компонента, контролирующая поведение, затрагивает регуляцию таких физиологических функций, как иммунитет, гемопоэз и активность пищеварительных ферментов.
Заключение: Изменения, происходящие в результате отбора по оборонительной реакции на человека, трудно интерпретировать с позиций мутационных изменений отдельных генов, контролирующих те или иные признаки. Сложность интерпретации формообразовательных процессов во взаимоотношении «человек–животное» отражает изменения в процессе отбора регуляторных генов, вовлекаемых в регуляцию генетической активности на разных стадиях морфогенеза, а потому играющих большую роль в эволюционных преобразованиях.
Финансирование: Работа поддержана бюджетным проектом Института цитологии и генетики СО РАН № FWNR-2022-0023.

Consequences of 26 generations of selection in minks (*Neogale vison*) on defensive reaction towards man

Trapezov O.V.[1, 2]*, Nekrasova M.A.[1], Stepanova M.A.[1, 3], Baishnikova I.V.[4], Ilyukha V.A.[5], Kalinina S.N.[4], Kizhina A.G.[4], Morozov A.V.[4], Panova E.V.[4], Balan O.V.[4]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Novosibirsk State Agrarian University, Novosibirsk, Russia*
[4] *Institute of Biology of Karelian Research Centre RAS, Petrozavodsk, Russia*
[5] *Papanin Institute for Biology of Inland Waters, RAS, Borok, Yaroslavl region, Russia*
* *trapezov@bionet.nsc.ru*

Key words: selection; *Neogale vison*; defensive behaviour; digestion; physiological traits

Motivation and Aim: During a long-term experiment on the selective transformation of the defensive reaction towards man in American minks, it was shown that selection opens up new opportunities for hereditary variability in the same direction in which selection itself acts. such selection is capable of creating an increase in the rate and scope of variability not only in the direction of its action, but also in other directions [Belyaev, 1968, 1970, 1972, 1974, 1979, 1981, 1983; Belyaev, Trut, 1964, 1982, 1983; Trut, 1999, 2007, 2008; Trapezov, 1987, 2012]. The aim of the work is to compare the rate and scope of variability in two lines of minks selected for tame and aggressive behavior.
Methods and Algorithms: Over 26 generations of selection in the American mink population of cage breeding of the *Standard dark brown* (+/+) genotype, two lines of animals with tame and aggressive behavior towards man were created. During the selection transformation of behavior, the following were calculated: (1) selection shift – the difference between the average values of a trait in the offspring of selected parents and in the parent generation as a whole; (2) selection differential – the difference

between the average values of a trait in individuals selected as parents of the next generation and in the whole population; 3) the value of realized heritability – h^2, to what extent the value of the selected trait is due to heredity and to what extent to the influence of the external environment; (3) the value of realized heritability – h^2, to what extent the value of the selected trait is due to heredity and to what extent to the influence of the external environment; (4) selection coefficients – *S* (%); (5) population inbreeding coefficients according to Falconer – *F*t; (6) actual responses to selection (*R*) per generation.

Results: During the experiment on the selective transformation of the defensive reaction, the relationship between the direction of selection and variability in behavior was identified. First of all, as can be seen in Figure 1, phenotypic variability during selection for domestication behavior is two orders of magnitude higher than in the group selected for aggressive behavior. Calculations of the average actual response to selection per generation show the greatest effect of selection on domesticated animals. Moreover, sexual dimorphism is observed in them – the greatest selection effect was recorded in males. In the expressiveness of tame behavior (males are significantly more tame than females ($p < 0.001$). Moreover, the value of the phenotypic variation of the selected trait in the domesticated population is two orders of magnitude higher than in the population selected for aggressiveness. The average actual response to selection for one The generation of aggressive animals, both males and females, is almost the same.

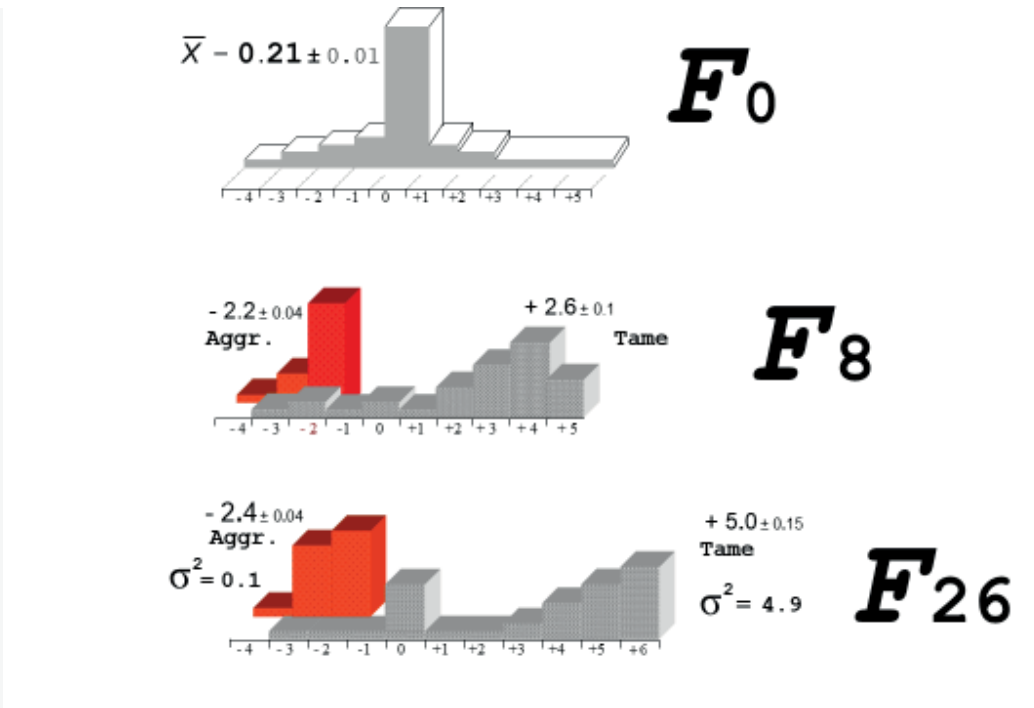


Fig. 1

Realized heritability, estimated as the ratio of the selection shift (R) to the selection differential (S), showed that about 60 % of the additive diversity is responsible for the response to the first two generations of “minus” selection, and about 40 % of “plus” selection (37 % for females and 38 % for males). The significant shift obtained during the first two generations of selection for aggressiveness indicates that the number of

main genes that determine the threshold of aggressive reactions is minimal, perhaps no more than two or three, and they are quickly fixed in the first stages of selection. Selection for a defensive reaction to humans leads to changes in the regulation of physiological functions. Animals from the aggressive line were characterized by higher basal levels of cortisol in the blood serum compared to tame animals. At the same time, tame individuals differed from aggressive ones by an increased content of eosinophils, but a decreased content of neutrophils, larger lymphocyte sizes and a lower nuclear-cell ratio. Tame minks showed an amylolytic profile of digestive enzyme activity and a higher content of vitamins A and E in the liver, while aggressive individuals had a higher activity of proteolytic and lipolytic enzymes of the digestive tract. It can be assumed that the gene component that controls behavior affects the regulation of such physiological functions as immunity, hematopoiesis and the activity of digestive enzymes.

Conclusion: Changes occurring as a result of selection for a defensive reaction towards man are difficult to interpret from the standpoint of mutational changes in individual genes that control certain traits. The complexity of interpreting the morphogenesis processes in the "human–animal" relationship reflects changes in the process of selection of regulatory genes involved in the regulation of genetic activity at different stages of morphogenesis, and therefore playing a large role in evolutionary transformations.

Funding: The study is supported by Institute of Cytology and Genetics, SB RAS No. FWNR-2022-0023.

doi 10.18699/bgrs2024-5.1-12

Генетическая гетерогенность восточной ночницы (*Myotis petax*, Hollister 1912)

Шаяхметова Л.[1, 2]*, Батурина А.[3], Маслов А.[4], Графодатский А.[1], Проскурякова А.[1]

[1] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

[4] *Институт систематики и экологии животных СО РАН, Новосибирск, Россия*

* *l.shayakhmetova@g.nsu.ru*

Ключевые слова: *Myotis petax*; митохондриальная ДНК; генетическое разнообразие; филогенетическое древо; гаплотипическая сеть

Мотивация и цель: Летучие мыши рода *Myotis* представляют собой богатый видами таксон летучих мышей и имеют практически глобальное распространение. Обитающая в Западной Сибири *Myotis petax* Hollister, 1912, характеризуется широким ареалом, включающим островные популяции, и высокой плотностью на зимовках. Кроме того, *Myotis petax* демонстрирует роение – поведение, при котором большие многовидовые скопления собираются для спаривания. Это явление наблюдается в период с августа по октябрь, когда каждую ночь у входов в пещеры собираются тысячи летучих мышей разных родов [1]. Данные особенности делают *Myotis petax* интересным объектом с точки зрения изучения внутривидовой изменчивости. Ранее были проведены работы по описанию генетического разнообразия этого вида на территории Восточной Сибири [2, 3]. Исследования генетического разнообразия *Myotis petax* на территории Западной Сибири позволяют оценить, насколько вид массовый и процветающий на данной территории.

Методы и алгоритмы: Была получена выборка, состоящая из 65 образцов *Myotis petax*, собранных на территории Западной Сибири (Новосибирская и Кемеровская области, Республика Алтай). Для образцов была выделена ДНК и методом ПЦР получены последовательности контрольного участка D петли митохондриальной ДНК [4]. На следующем этапе мы провели филогенетический анализ, основанный на выравнивании нуклеотидных последовательностей участка мтДНК в программе MEGA [5]. Для анализа дополнительно были включены последовательности контрольного региона 32 особей, отловленных в разных географических регионах, из базы данных GenBank. Филогенетические отношения были реконструированы с помощью метода максимального правдоподобия (ML) [5]. В качестве внешней группы были выбраны 6 образцов из Кореи и Китая. Построение гаплотипической сети осуществлено с использованием программы POPART [6].

Результаты: В результате работы была построена гаплотипическая сеть, которая отличалась звездчатым типом. На территории Западной Сибири и Дальнего Востока популяции *Myotis petax* характеризуются высоким генетическим разнообразием. На территории от Новосибирской области и до Дальнего Востока не наблюдается кластеризации отдельных групп по географическим характеристикам, поэтому мы можем предположить, что внутри вида восточной

ночницы отсутствуют репродуктивные преграды. Образцы из Кореи образовали сильно поддерживаемую кладу, более отдаленную, чем дальневосточные. Это может свидетельствовать о том, что существует поток генов между Западной Сибирью и Дальним Востоком, при этом популяция из Кореи имеет тенденцию к большей изолированности.
Выводы: Звездчатая структура гаплотипической сети свидетельствует о высоком уровне генетического полиморфизма у восточных ночниц на территории Западной Сибири и может указывать на то, что в недавней истории вида *Myotis petax* не происходило резкого сокращения численности и снижения разнообразия.
Финансирование: Исследование поддержано грантом РНФ (№ 19-14-00034-П).

Genetic heterogeneity of *Myotis petax*, Hollister 1912

Shayakhmetova L.[1, 2]*, Baturina A.[3], Maslov A.[4], Graphodatsky A.[1], Proskuryakova A.[1]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Lomonosov Moscow State University, Moscow, Russia*

[4] *Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia*

* *l.shayakhmetova@g.nsu.ru*

Key words: *Myotis petax*; mitochondrial DNA; genetic diversity; phylogenetic tree; haplotype network

Motivation and Aim: *Myotis* bats comprise the most species-rich bat genus and have an almost global distribution. *Myotis petax* Hollister, 1912, living in Siberia, is characterized by a wide habitat, including island populations, and high density in hibernation sites. Also, *Myotis petax* exhibit autumnal swarming, a behavior where large, multi-species assemblages gather for mating. This phenomenon occurs between August and October, with thousands of bats from various genera gathering at cave entrances each night [1]. These characteristics make *Myotis petax* a fascinating subject for exploring intraspecific variability. Previously, work was carried out to describe the genetic diversity of this species in Eastern Siberia [2, 3]. Studies of the genetic diversity of *Myotis petax* in Western Siberia make it possible to assess how widespread and flourishing the species is in this territory.
Methods and Algorithms: During the study, 65 samples of *Myotis petax* were collected in Western Siberia (Novosibirsk and Kemerovo regions, Altai Republic). DNA was isolated for the samples and the sequences of the control region D loop of mitochondrial DNA were obtained by PCR [4]. At the next stage, we carried out a phylogenetic analysis based on the alignment of nucleotide sequences of the mtDNA region in the MEGA program. The control region sequences of 32 individuals from different geographic regions sourced from the GenBank database were additionally included for analysis. Phylogenetic relationships were reconstructed using the maximum likelihood (ML) method [5]. Additionally, 6 samples from Korea and China were selected as the outgroup. The haplotype network was constructed using the POPART program [6].
Results: As a result of the work, a haplotype network was constructed, which was distinguished by a star type. In Western Siberia and the Far East populations of *Myotis petax* are characterized by a high genetic diversity. Over the territory from Novosibirsk to the Far East no clustering of individual groups according to geographical

characteristics has been observed. We can assume that there are no reproductive barriers within the species of *Myotis petax*. Samples from Korea formed a strongly supported clade, more distant than those from the Far East. This may indicate that there is a gene flow between Siberia and the Far East, with the population from Korea tending to be more isolated.

Conclusion: In general, the star-shaped structure of the haplotype network indicates a high level of genetic polymorphism in *Myotis petax* in Western Siberia and may indicate that in recent history of the *Myotis petax* species there hasn't been a sharp decline in number and diversity.

Funding: The study is supported by the Russian Science Foundation (No. 19-14-00034-П).

Список литературы/References

1. Foley N.M., Harris A.J., Bredemeyer K.R. et al. Karyotypic stasis and swarming influenced the evolution of viral tolerance in a species-rich bat radiation. *Cell Genomics*. 2024;4(2):100482. doi 10.1016/J.XGEN.2023.100482
2. Gorobeyko U.V., Kartavtseva I.V., Sheremetyeva I.N., Kazakov D.V., Guskov V.Y. DNA-barcoding and a new data about the karyotype of Myotis petax (Chiroptera, Vespertilionidae) in the Russian Far East. *Comp Cytogenet*. 2020;14(4):483-500. doi 10.3897/CompCytogen.v14i4.54955
3. Gorobeyko U.V. et al. A new type of tandem repeats in Myotis petax (Chiroptera, Vespertilionidae) mitochondrial control region. *Mol Biol Rep*. 2023;50(6):5137-5146. doi 10.1007/s11033-023-08468-4
4. Горобейко У.В. Морфрлогическая и генетическая изменчивость восточной ночицы Myotis petax Hollister, 1912 на юге Дальнего Востока России. Дисс. канд. биол. наук. 2021
[Gorobeyko U.V. Morphrological and genetic variability of the eastern noctule Myotis petax Hollister, 1912 in the south of the Russian Far East. Diss. Ph. D. biol. Sci. 2021 (in Russian)]
5. Tamura K., Stecher G., Kumar S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021;38(7):3022-3027. doi 10.1093/MOLBEV/MSAB120
6. Leigh J.W., Bryant D. POPART: full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6:1110-1116. doi 10.1111/2041-210X.12410

doi 10.18699/bgrs2024-5.1-13

Evolution of embryonic diapause in mammals and applying reproductive technologies to diapausing embryos

Amstislavsky S.[1]*, Rakhmanova T.[1, 2], Rozhkova I.[1], Kozeneva V.[1, 2], Okotrub S.[1], Lebedeva D.[1], Igonina T.[1], Babochkina T.[1], Omelchenko A.[2, 3], Okotrub K.[3], Brusentsev E.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Institute of Automation and Electrometry, SB RAS, Novosibirsk, Russia*

* *amstis@bionet.nsc.ru*

Key words: embryonic diapause; evolution; reproductive technologies; diapausing embryos; cryopreservation

Motivation and Aim: Arrests in the embryo development are characteristic to species belonging to various taxa of the animal kingdom [1]. Among mammals, obligate embryonic diapause (ED), i. e. suspension of embryo development at the blastocyst stage and delayed implantation, present in each development cycle, occurs in no less than 60 species belonging to 12 families [2]. Because of ED the periods of mating and the offspring birth may be attributed to the most favorable seasons [1, 2]. The evolutional aspects of ED in mammals are discussed in the literature, e. g. whether ED is a trait acquired during evolution *de novo* or is it characteristic of the ancestral forms and the loss of this trait occurs in some species [3–5]. Modern reproductive technologies, such as embryo *in vitro* culture (IVC), cryopreservation, and embryo transfer may be helpful in addressing issues related to the evolution of diapause and provide an insight to disclose the mechanism of this reproductive phenomenon [5–8]. In particular, by transferring ovine embryos from sheep into the reproductive tract of the mouse, it was possible to induce a reversible state of diapause, which sheep embryos normally lacking; this leads the authors to formulate the hypothesis that genes associated with diapause are evolutionarily conserved [5]. Experimental modeling of ED in farm and laboratory animals is valuable not only to address the questions related to the evolution of this trait, but also helpful for the Genome Resource Banking concept applying to diapausing mammalian species. As some rare and threatened mammals exhibit ED, there is motivation to develop reproductive technologies especially targeting diapausing embryos [9]. The goal of this work was to develop reproductive technologies for diapausing embryos using the mouse as an experimental model.

Methods and Algorithms: The studies were carried out on CD1 mice kept under standard conditions of SPF-Vivarium and approved by the Bioethics Committee of the Institute of Cytology and Genetics SB RAS (No. 144, 29.03.2023). Diapausing embryos were collected on the 6th day *post coitum* (dpc), after fertile mating of CD1 mice, treated with equine chorionic gonadotropin (eCG) and human chorionic gonadotropin (hCG). The mice were ovariectomized on the 4th dpc and received progesterone injections on the 4th and 5th dpc (diapause group) [6, 7]. Cryopreservation was performed using a CL-8800i programmable controlled rate freezer (CryoLogic, Australia) with 1.8 M ethylene glycol and 0.1 M sucrose as cryoprotectant agents [6, 7]. The viability of frozen-thawed diapausing embryos was evaluated *in vitro* by their culturing during 24 h and 47 h in the

CO_2-incubator New Brunswick™ Galaxy 48R CO_2-incubator at 37 °C, 5 % CO_2, and 90 % humidity in 20-μL drops of the KSOM (Merck, Germany) supplemented with 1 % essential amino acids, 0.5 % non-essential amino acids, 0.5 % vitamins, and 5 % fetal calf serum (Thermo Fisher Scientific, USA), covered with mineral oil (FertiPro, Belgium). The embryonic development was monitored by visual examination under an S8 APO microscope (Leica Microsystems, Germany). The development rate after 24 h IVC was estimated as the percentage of blastocysts with expanded blastocoel cavity from the total number of thawed blastocysts [6–8]. Randomly chosen frozen-thawed diapausing embryos were fixed immediately after thawing as well as after 24 h and 47 h of IVC and processed for DAPI/TUNEL staining to evaluate interphase cell numbers as well as apoptotic and fragmentation indexes as described earlier [6]. The results were analyzed using the STATISTICA v. 12.0 software (StatSoft, Inc., USA). Data were analyzed using analysis of variance (one-way ANOVA) followed by *post-hoc* comparison (Fisher LSD-test).

Results: The yield of embryos per diapausing mouse female was 9.6 ± 1.5 [6] and 10.4 ± 1.5 [7]. All the diapausing embryos collected were at blastocyst stage. Total number of diapausing embryos cryopreserved in two independent studies was 160 (11 replicates) [6, 7]. Diapausing murine blastocysts have already lost *zona pellucida* and are characterized by the large size as the volume of the blastocoel increases during the period of implantation delay (Fig. 1). Some other mammalian species are characterized by even larger blastocysts during diapause [10]. This makes cryopreservation of diapausing embryos challenging. However, the percentage of blastocysts with expanded blastocoel cavity after freezing-thawing and subsequent *in vitro* culture was relatively high: 82.5 % [6] and 85.0 % [7]. Moreover, the viability of frozen-thawed blastocysts was confirmed by different methods. Cryopreservation did not cause a decrease of protein metabolism in diapausing embryos [6]. In addition, cryopreservation did not cause a decrease of *Odc1* and *RhoA* gene expression in diapausing embryos [7].

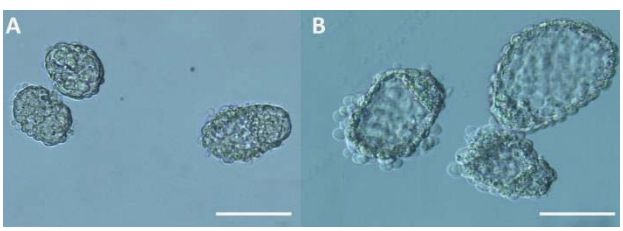


Fig. 1. Diapausing murine embryos after cryopreservation. A, diapausing frozen-thawed blastocysts before IVC, collapsed blastocoel; B, diapausing blastocysts with re-expansion of blastocoel, 24 h IVC after thawing. Scale bar 100 μm

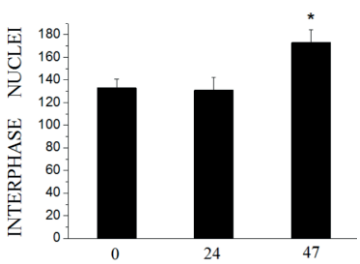


Fig. 2. Cell numbers (M ± SEM) in diapausing murine embryos after freezing-thawing and subsequent IVC in a medium supplemented with 100 μM putrescine. * $P < 0.05$

The increase of cells in the frozen-thawed embryos after IVC in a medium containing putrescine is another evidence of their viability (Fig. 2). The analysis using ANOVA revealed a significant effect of the culturing time on the number of interphase nuclei per embryo [$F(2,21) = 4.9$, $P < 0.05$].

Conclusion: Embryonic diapause disentangles breeding season and the offspring birth, thus evolutionally perpetuate the species. We studied the possibility of diapausing embryos program freezing using the mouse as a model species. The post-thaw embryo viability was assessed by different criteria; all these criteria indicated that the diapausing murine embryos survived cryopreservation. Meanwhile, our study revealed that diapausing embryos need re-activation time to resume cell division after freezing-thawing procedures. To expand embryo cryopreservation technology to other mammalian species, which possess delayed implantation, the evolutional difference in diapause phenotype between species of mammals should be taken into account.

Funding: The study is supported by the Russian Science Foundation (Project No. 23-24-00313).

Список литературы/References

1. Easwaran S., Montell D. The molecular mechanisms of diapause and diapause-like reversible arrest. *Biochem Soc Trans*. 2023;51(5):1847-1856. doi 10.1042/BST20221431
2. Fenelon J.C. et al. Embryonic diapause: development on hold. *Int J Dev Biol.* 2014;58:163-174. doi 10.1387/ijdb.140074bm
3. Thom M. et al. The evolution and maintenance of delayed implantation in the Mustelidae (Mammalia: Carnivora). *Evolution.* 2004;58(1):175-183. doi 10.1111/j.0014-3820.2004.tb01584.x
4. Lindenfors P. et al. The monophyletic origin of delayed implantation in carnivores and its implications. *Evolution.* 2003;57(8):1952-1056. doi 10.1111/j.0014-3820.2003.tb00601.x
5. Ptak G. et al. Embryonic diapause is conserved across mammals. *PLoS One.* 2012;7:e33027. doi 10.1371/journal.pone.0033027
6. Amstislavsky S. et al. Program freezing of diapausing embryos in the mouse. *Theriogenology.* 2024;217:1-10. doi 10.1016/j.theriogenology.2024.01.006
7. Amstislavsky S. et al. Effect of cryopreservation on Odc1 and RhoA genes expression in diapausing mouse blastocysts. *Reprod Domest Anim.* 2024:59(5):e14576. doi 10.1111/rda.14576
8. Wang X. et al. Cathepsin L involved in the freezing resistance of murine normal hatching embryos and dormant embryos. *Reprod. Biol.* 2022;22:100612. doi 10.1016/j.repbio.2022.100612
9. Wauters J. et al. Could embryonic diapause facilitate conservation of endangered species? *Bioscientifica Proc*. 2020;10:76-84. doi 10.1530/biosciprocs.10.005
10. Amstislavsky S., Brusentsev E., Kizilova E. Delayed implantation combined with precocious sexual maturation in female offspring: a story of the stoat. *Bioscientifica Proc*. 2020;10:197-206. doi 10.1530/biosciprocs.10.014

doi 10.18699/bgrs2024-5.1-14

A principal components analysis and functional annotation of differentially expressed genes in brain regions of gray rats selected for tame or aggressive behavior

Chadaeva I.[1]*, Kozhemyakina R.[1], Shikhevich S[1], Bogomolov A.[1, 2], Kondratyuk E.[1, 3, 4], Oshchepkov D.[1, 2], Orlov Yu. L.[5, 6]*, Markel A.L.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia;*

[2] *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

[3] *Siberian Federal Scientific Centre of Agro-BioTechnologies, Krasnoobsk, Russia*

[4] *Research Institute of Clinical and Experimental Lymphology, Novosibirsk, Russia*

[5] *Institute of Biodesign and Complex Systems Modeling, Moscow, Russia*

[6] *Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia*

** ichadaeva@bionet.nsc.ru*

Key words: animal domestication; artificial selection; behavioral genetics; brain; differentially expressed gene; aggressive and tame rats; RNA-seq; principal component analysis; functional annotation of genes; gene network

Motivation and Aim: The aim of this work was to analyze and carry out functional annotation of differentially expressed genes (DEG) in four brain regions (the hypothalamus, the hippocampus, the periaqueductal gray matter (PAG), and the tegmental region of the midbrain (MTg) of gray rats selected for tame or aggressive behavior.

Methods and Algorithms: To analyze the DEGs, we used our previously obtained high-throughput RNA sequencing data from four brain regions of gray rats of the two strains: the hypothalamus [1], the hippocampus [2], the PAG [3], and the MTg [4]; these data have been deposited in the NCBI SRA database (ID = PRJNA668014). The total list of DEGs was subjected to principal component analysis: we performed this analysis using the PAST4.04 software package. Functional annotation of DEGs was carried out with the help of Web service DAVID Bioinformatics Resources. The construction of gene networks based on interactions between the DEGs under study was performed by means of web service STRING.

Results: On the basis of previously published transcriptomes of the hypothalamus, hippocampus, PAG, and MTg of tame and aggressive rats, we compiled a total list of DEGs, which consisted of 112 genes. The data on differential gene expression were then subjected to principal component analysis (PCA). It revealed a first principal component (here and further: PC1), which explained the largest proportion of variance of gene expression in the brain regions between the two strains (29 %) and a second principal component (here and further: PC2; 17 % of the variance). As for the PC1, the rats were subdivided into two groups on the basis of the 26 genes (out of 112 DEGs) that showed a significant correlation with this principal component (Fig. 1).

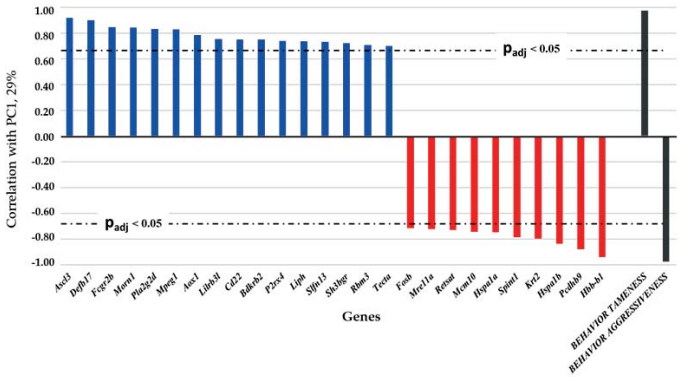


Fig. 1. A statistically significant linear correlation between expression levels of DEGs and the PC1. Genes whose expression levels are higher in tame rats are highlighted in blue; the genes whose expression is higher in aggressive rats are highlighted in red

The PC1 reflects gene expression changes that are caused by artificial selection based on the behavioral response of rats to humans. Accordingly, this component's positive coefficient of correlation with 16 genes points to a connection with the tame behavioral phenotype, and the negative correlation with 10 genes indicates an association with the rat phenotype of the aggressive response. The contribution of the PC2 accounted for only 17 % of the total variance of gene expression under our experimental conditions. The PC2 correlates with the expression levels of five genes in the hippocampus and reflects hippocampus-specific changes of gene expression in tame and aggressive rats.
Functional annotation of the DEG list was carried out using DAVID (Fig. 2).

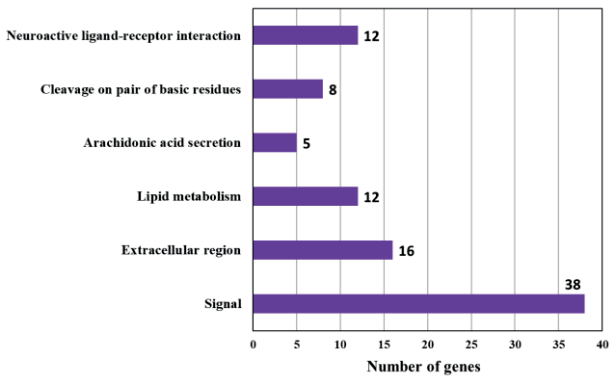


Fig. 2. The distribution of all DEGs among GO and KEGG terms in the brain regions of tame and aggressive rats according to DAVID. Biological processes are presented that are significantly enriched within the total list of DEGs

The DAVID analysis made it possible to determine which biological processes could lead to changes in tame and aggressive rats: "Neuroactive ligand-receptor interaction", "Cleavage on pair of basic residues", "Arachidonic acid secretion", "Lipid metabolism", "Extracellular region", and "Signal."
To identify possible relations between some DEGs a gene association network was built using STRING (Fig. 3). The gene network contains one large cluster, which includes such relevant genes as *Pcp2*, *Nmb*, *Fosb*, *Nr4a3*, *Shox2*, *Hspa1a*, *Hspa1b* and others: a total of 77 DEGs out of 112. In this total list, 33 DEGs were found to not participate in any interactions.

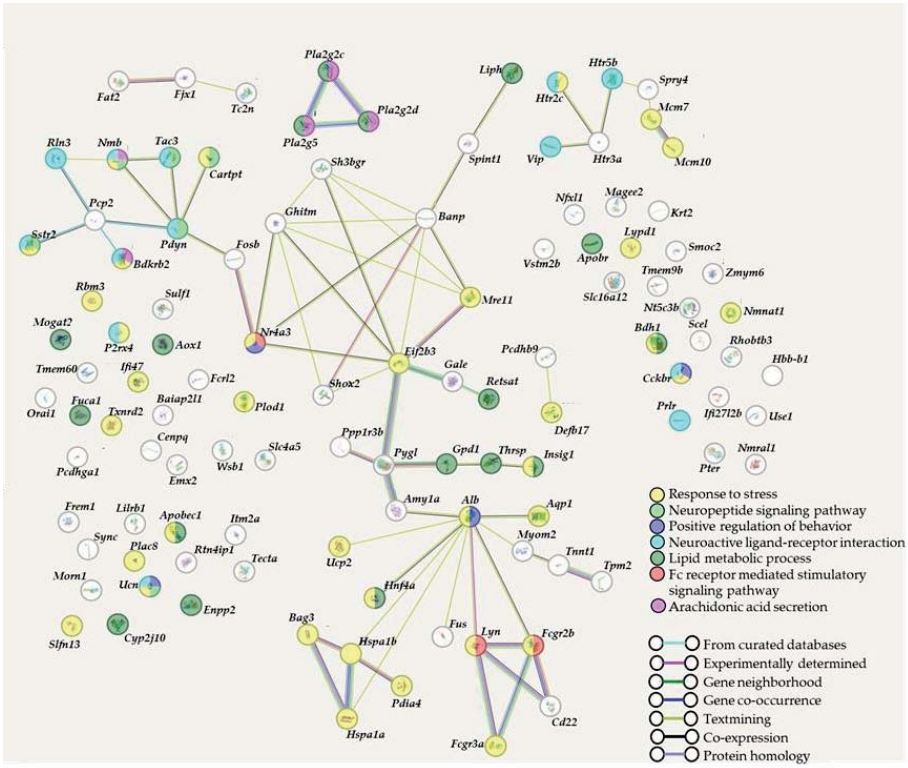


Fig. 3. The gene association network built by means of web service STRING on the basis of protein–protein interactions corresponding to some DEGs of tame and aggressive rats. Genes that are related to one of seven biological processes are highlighted in a color(s), and interactions between genes are indicated by lines

Thus, the DAVID and STRING functional enrichment analysis results complement each other and identify overrepresented terms that drive behavioral responses. All this together determines a multifactor system controlling the two different types of behavior towards human for which selection is carried out.

Conclusion: We studied the transcriptomes of two strains of gray rats selected for behavior towards human. We identified and analyzed genes that change their expression in rat brain regions as a result of the artificial selection for tame or aggressive behavior. The results obtained confirm the presence of a multifactorial system for controlling the behaviors studied.

Funding: The study is supported by Russian government projects FWNR-2022-0015.

References

1. Chadaeva I. et al. Domestication explains two-thirds of differential gene expression variance between domestic and wild animals; the remaining one-third reflects intraspecific and interspecific variation. *Animals.* 2021;11:2667. doi 10.3390/ani11092667
2. Oshchepkov D. et al. Stress reactivity, susceptibility to hypertension, and differential expression of genes in hypertensive compared to normotensive patients. *Int J Mol Sci.* 2022;23(5):2835. doi 10.3390/ijms23052835
3. Shikhevich S. et al. Differentially expressed genes and molecular susceptibility to human age-related diseases. *Int J Mol Sci* 2023;24(4):3996. doi 10.3390/ijms24043996
4. Oshchepkov D. et al. Transcription factors as important regulators of changes in behavior through domestication of gray rats: quantitative data from RNA sequencing. *Int J Mol Sci.* 2022;23(20):12269. doi 10.3390/ijms232012269

doi 10.18699/bgrs2024-5.1-15

Hydrophobic and aerophobic mitochondrial mutational spectrum of birds

Gusarov Y.[1]*, Mikhailova A.[1], Efimenko B.[1], Gunbin K.[1], Bushuev A.[2], Burskaya V.[3], Popadin K.[1]

[1] *Center of the Genomic Researches, Higher School of Living Systems, Institute of Medicine and Life Science, Immanuel Kant Baltic Federal University, Kaliningrad, Russia*

[2] *Lomonosov Moscow State University, Moscow, Russia*

[3] *University of Antwerp, Antwerp, Belgium*

* *yurgusguss@mail.ru*

Key words: mitochondria; mtDNA; evolution; mutational spectrum; birds

Motivation and Aim: Mitochondria is the powerbase of the cells with its own DNA molecule (further mtDNA). During replication mtDNA becomes vulnerable to different mutagens. Replication in mtDNA strands proceeds unevenly [1]. Protein-coding genes, close to the origin of replication (COX1, COX2) spend less time in single-stranded condition than ones farther from it (CytB) on heavy strand [2]. Exposed genes in single stranded states become easy targets for different internal mutagenic factors, including spontaneous deamination or even oxidative damage. The result of this mutagenic process is A>G substitution on the heavy strand. It was shown for different groups of animals that this mutation is connected with different life-history traits. Long-lived mammals demonstrate increased frequencies of A>G transitions compared to the short-lived ones [3]. Also there is a high difference in mutation frequencies of A>G between fishes inhabiting cold and warm water [4]. Armed by this knowledge we decided to analyse mtDNA mutagenesis in birds and potential factors affecting it.

Methods and Algorithms: To study the influence of life-history traits in different birds on their mtDNA mutational spectrum, we collected all available information. First of all we analysed mtDNA RefSeq data for 766 bird species. For each species we calculated a specific metric to study mtDNA nucleotide shift due to A>G mutation – GhAhSkew by using the following formula: GhAhSkew = (G-A)/G+A). In this formula G and A is the amount of guanine and adenine in four-fold neutral positions. We used it in each analysis to study mtDNA mutagenesis. Moreover we added ecological data from the AVONET database and deployed a specific phenotype classification of birds based on ability to fly and dive with the help of Birds of the World database [5, 6]. To find primary associations between mtDNA mutagenesis and life-history traits we used t-test, PCA, U-test and linear models. To study phylogenetic signals in the observed associations we used specific methods such as PGLS and reconstruction of ancestral states.

Results: By using simple statistical methods we discovered pretty interesting results for RefSeq data. Similar to mammals, birds have a gradient of GhAhSkew in protein-coding genes depending on their time being single stranded during replication [3]. COX1 and COX2 have low GhAhSkew compared to CytB. Having such results we decided to compare each group of animals. We saw that birds have around 1.5 times higher GhAhSkew than mammals. This result drew us to a thought that birds' mutagenesis speed was much higher than mammals. In order to understand the reasons behind such intense mutagenesis, we started to compare our metric GhAhSkew between different birds' life-

history traits. We observed that birds with only one trait had low GhAhSkew – losing ability to fly. Primarily for *Palaeognathae* species. But there were many traits of birds which increased GhAhSkew: ability to dive (especially for *Sphenisciformes* and *Anseriformes*), far migration, thermo-neutral zone. To our surprise, there was no connection between GhAhSkew and many metrics, like body mass for example. After all these discoveries we decided to analyse the phylogenetic value of them. Due to the data limits we were only able to look at ability to fly/dive phenotypes. For all those phenotypes PGLS showed statistically significant results for GhAhSkew and losing ability to fly (p-value < 0.05) and GhAhSkew and getting ability to dive (p-value < 0.05). Both results also had high Pagel's lambda value ($\lambda > 0.9$) which shows strong phylogenetic inertia for the GhAhSkew metric. For those phenotypes we also calculated PGLS with a comparative species-specific A>G spectrum, which is based on our RefSeq data. We found no correlations for A>G frequencies and any phenotypes. We made a hypothesis that birds with our phenotypes of interest are under stabilizing selection in order to save last adenines.

Conclusion: We showed that birds mtDNA A>G mutation rate is quite high, even compared to mammals. It led to increased guanine enrichment in modern birds mtDNA with specific phenotypes: ability to dive, long-distance migration and thermo-neutral zone. But there are also factors for lowering mutagenesis – losing ability to fly. We discovered that all mutagenesis effects emerged at the level of ancestors, and nowadays some bird species are under stabilizing selection. For further research, we have expanded our data by gathering mtDNA information from the MIDORI2 database for more than 7000 bird species [7]. In addition we also got A>G spectra for our birds database: one for our 766 RefSeqs and another for 789 birds by using the NeMu pipeline [8].

Funding: The study is supported by RSF grant (No. 21-75-20143).

References

1. Falkenberg M. Mitochondrial DNA replication in mammalian cells: overview of the pathway. *Essays Biochem*. 2018;62(3):287-296. doi 10.1042/EBC20170100
2. Ju Y.S., Alexandrov L.B., Gerstung M. et al. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *Elife*. 2014;3:e02935. doi 10.7554/eLife.02935
3. Mikhailova A.G., Mikhailova A.A., Ushakova K. et al. A mitochondria-specific mutational signature of aging: increased rate of A > G substitutions on the heavy strand. *Nucleic Acids Res*. 2022;50(18):10264-10277. doi 10.1093/nar/gkac779
4. Mikhailova A. et al. A mitochondrial mutational signature of temperature in ectothermic and endothermic vertebrates. *bioRxiv*. 2021. doi 10.1101/2020.07.25.221184
5. Tobias J.A., Sheard C., Pigot A.L. et al. AVONET: morphological, ecological and geographical data for all birds. *Ecol Lett*. 2022;25:581-597
6. Billerman S.M., Keeney B.K., Rodewald P.G., Schulenberg T.S. (Eds.). Birds of the World. Cornell NY, USA: Laboratory of Ornithology, Ithaca, 2022
7. Leray M., Knowlton N., Machida R.J. MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences. *Environmental DNA*. 2022;4(4):894-907. doi 10.1002/edn3.303
8. Efimenko B., Popadin K., Gunbin K. NeMu: a comprehensive pipeline for accurate reconstruction of neutral mutation spectra from evolutionary data. *bioRxiv* 2023. doi 10.1101/2023.12.13.571433

doi 10.18699/bgrs2024-5.1-17

Possible genetic isolation between different sympatric morphs of Dolly Varden (*Salvelinus malma*) from Kronotskoe lake

Ignatiev B.D., Scobeyeva V.A.*

Department of Biological Evolution, Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia

** skobei-khanum@yandex.ru*

Key words: Dolly Varden; sympatric morphs; transcroptome

Motivation and Aim: Arctic chars (Salvelinus alpinus) and Dolly Varden (Salvelinus malma) are wide-spead circumpolar specia of Salmonids. They produce many sympatric and allopatric morphs with unclear taxonomic status and even sometimes considered one species. Dolly Warden inhabits rivers and lakes of Pacific basin, and were so we can speak about fish from Kronotskoe Lake as Dolly Varden, taking into account data on mtDNA [1]. Kronotskoe lake is inhabited by multiple sympatric morphs of Dolly Varden, some authors define to 8 [2]. In this study we make an attempt to solve the problem of genetic isolation between sympatric Dolly Varden morphs in Kronotskoe lake using morph transcriptomic data.

Methods and Algorithms: Sampes of muscle tissue were taken from fish, caught in Kronotskoe lake in 2013. Samples were stored in RNAlater solution and transported to University of Southern California, then libraries were prepared and Illumina sequencing was performed. 150-bp pair end reads were obtained for each fish and trimmed by Trimmomatic. Then reads were aligned with reference genome of *Salvelinus* sp., putative hybrid between *S. alpinus* and *S. malma* [3]. Assembly ASM291031v2, RefSeq GCF_002910315.2) with hisat2 (Kim et al., 2019; RRID:SCR_015530 [4]). The bam files for each fish were generated with samtools and polymorphisms were searched by bcftools. Annotation of SNP was made with snpEff и snpSift.

Results: Transcriptomic reads were successfully mapped on reference genome (average 89.29 % reads mapped) 1697 SNP were found and annotated, most of them were modifiers (84.5 %) and silent (82.78 %). The most abundant locations were downstream (38.18 %) and upstream (33.33 %) variants. Most SNP were silent (82.7 %), less abundant were missense (16 %) and nonsense (0.566 %) mutations. PCA analysis of all SNPs revealed low portion of variance explained by first two components. PC1 for the whole list of SNPs was 5.32 %, PC2 – 13.50 %. Taken together PC1 and PC2 explain 28.82 % of variance. For mitDNA PC1 explains 27.82 % of variance, PC2 – 20.37 %, taken together PC1 and PC2 explain 48.19 %. For moprh-specific SNPs were determined GO. Dwarfs had the highest proportion of unique SNPs among all morphs and the highes percent of SNPs with high and low impact on affected gene (0.791 and 20.261 %). Also dwarf morph has the lowest percent of missense mutations and highest of silent (27.916 and 71.409 %). Fst analysis revealed no significant genetic distance between mophs, we can speak only of mild reproductive isolation between dwarf morph and other (Fst = 0.1 between dwarf and long-head morphs).

Conclusion: We found no evidence of reproductive isolation between different Dolly Varden morphs in Kronotskoe lake. Only mild traces of genetic isolation can be found in the dwarf morph, probably due to its longer history. We can suppose than dwarf morph

was lake-spawning non-migrating population before Kronotskoe was locked by volcanic eruption.

References

1. Weinstein S.Y., Gallagher C.P., Hale M.C. et al. An updated review of the post-glacial history, ecology, and diversity of Arctic char (Salvelinus alpinus) and Dolly Varden (S. malma). *Environ Biol Fish*. 2024;107:121-154. doi 10.1007/s10641-023-01492-0
2. Esin E.V., Markevich G.N. Evolution of the charrs, genus Salvelinus (Salmonidae). 1. Origins and Expansion of the Species. *J Ichthyol*. 2018;58(2):187-203. doi 10.1134/S003294521802005
3. Christensen K.A., Rondeau E.B., Minkley D.R. et al. The Arctic charr (Salvelinus alpinus) genome and transcriptome assembly. *Plos One*. 2018;13(9):e0204076. doi 10.1371/journal.pone.0204076
4. Kim D., Paggi J.M., Park C. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37:907-915. doi.org/10.1038/s41587-019-0201-4

doi 10.18699/bgrs2024-5.1-18

Association of SNP in the *KAT6B* gene with live weight of heifers of the Yaroslavl cattle breed

Igoshin A.V.[1], Yudin N.S.[1, 2], Ilina A.V.[3], Larkin D.M.[4]*

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Federal Williams Research Center for Forage Production and Agroecology, Scientific Research Institute of Livestock Breeding and Forage Production, Yaroslavl Region, Russia*

[4] *Royal Veterinary College, University of London, London, United Kingdom*

* *dmlarkin@gmail.com*

Key words: cow; Yaroslavl breed; weight; withers height; *MSS51* gene; *KAT6B* gene; single nucleotide polymorphism; association

Motivation and Aim: Previously, using the whole genome genotyping, we showed that the Yaroslavl breed has unique genetics compared to other Russian and foreign cattle breeds, while foreign breeds had an insignificant impact on the gene pool of Yaroslavl cattle [1]. Analysis of whole genome sequencing data of the Yaroslavl cows and data from the "1000 Bull Genomes" project made it possible to identify a unique haplotype on chromosome 28, which is found almost exclusively in Yaroslavl cattle breed [2]. This haplotype contains the *KAT6B* gene and appears to be negatively associated with withers height, heart girth and live weight at different ages. So probably, it has undergone purifying selection in other cattle breeds.

Methods and Algorithms: DNA fragment containing the missense variant G>A (p.Val105Met) within the *KAT6B* gene was amplified by PCR. The amplicons were cut with *HpySE526 I* restriction enzyme (SibEnzyme Ltd). The association between phenotypes obtained from individual heifer record cards and genotypes was estimated using linear regression ('lm' R function).

Results: We genotyped an additional 112 heifers to study the negative association of the Val105Met mutation in the *KAT6B* gene with live weight in Yaroslavl cows. The association we found earlier was confirmed in a pooled sample of 142 animals (Table 1, Fig. 1).

Table 1. Associations of Val105Met mutation in the *KAT6B* gene with weight of heifers of the Yaroslavl cattle breed at different ages. LM is a linear model, LM-PCs is a linear principal component-corrected model, and β is the effect size (kg/allele)

Weight	LM		LM-PCs	
	β	*p*	β	*p*
At birth	–0.28	0.5828	–0.27	0.5732
6 months	–6.76	0.0036	–6.25	0.0054
10 months	–5.68	0.0667	–5.77	0.0334
12 months	–8.36	0.0219	–8.2	0.0103
18 months	–12.62	0.0078	–11.25	0.0093

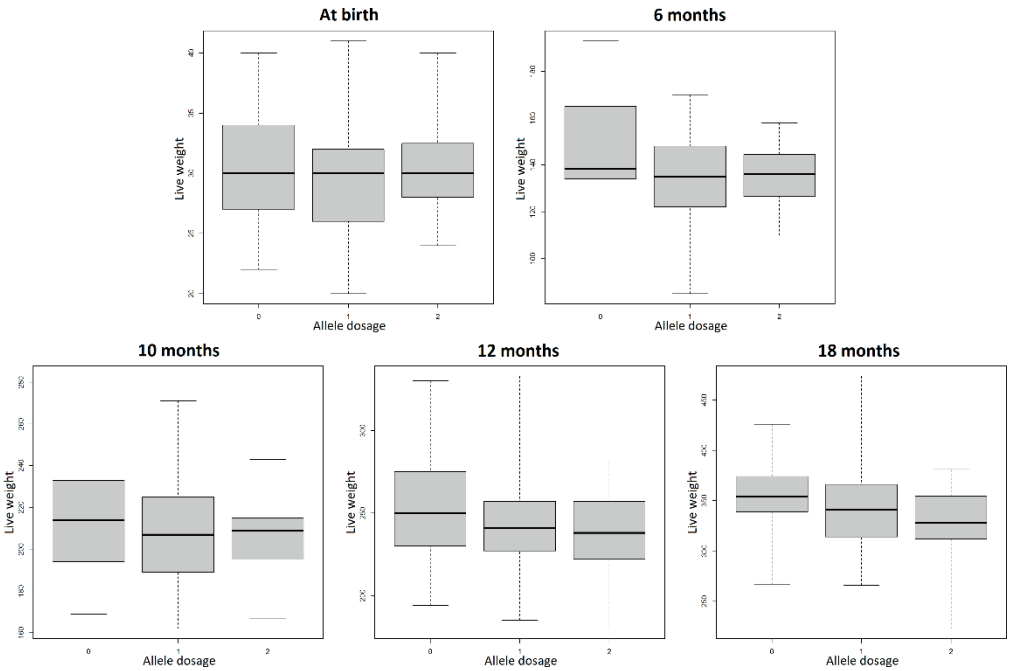


Fig. 1. Box and whisker plots illustrating the association between the dose of the methionine allele and the live weight of heifers of the Yaroslavl cattle breed at different ages

At the same time, because the system of reproduction in cattle is significantly different from panmixia, it can be assumed that the observed associations may be due to an unaccounted factor of the origin of animals. We conducted an additional analysis using pedigree data from animal records to verify this assumption. Using the "kinship" function of the R "kinship2" library, a kinship matrix 142 × 142 was constructed based on a genealogy spanning two generations (paternal and maternal grandparents). Then the principal component analysis was performed using the standard "eigen" R-function (Fig. 2).

We analyzed the association between the principal components and the live weight of animals at different ages. Overall, the strongest association with weight was found for the second component (PC2). It was statistically significant from 6 to 18 months. The association was strongest at the age of 18 months (p=1.80E-06). This is probably due to the fact that the PC2 component allows us to distinguish a cluster of animals that are descendants of the bull Marshal-1073 (Fig. 2), which apparently has a low breeding value for live weight, because 15 of its 16 daughters at the age of 18 months have a weight less than the sample average. Association in at least one of the ages was also observed for the 3rd, 4th, 5th, 8th, 10th, 15th and 18th components considering those accounting for more than 1 % of the variance explained. These components explain ~17 % of the total additive genetic variance. We used these components as covariates in the linear model in order to account for the effect of animal kinship in the association analysis.

As a result, the previously significant ($p < 0.05$) associations were retained. Moreover, the trend level association ($p < 0.10$) observed at 10 months of age reached a significant level after adjusting for principal components (Table 1). This suggests that the observed association does not represent a false-positive result due to the hidden effects of animal relatedness or sample stratification.

Fig. 2. Population structure of the analyzed sample visualized using two principal components. An oval in the lower left corner indicates a cluster of descendants of the bull Marshal-1073

Conclusion: The Yaroslavl cattle-specific Val105Met variant in the *KAT6B* gene is negatively associated with live weight of heifers at different ages. A further study exploring the relation of this mutation to other economically important phenotypes will elucidate its potential for use in marker-assisted selection.

Funding: The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia) according to the agreement with the Ministry of Education and Science RF, No. 075-15-2019-1662.

References

1. Yurchenko A., Yudin N. et al. Genome-wide genotyping uncovers genetic profiles and history of the Russian cattle breeds. *Heredity (Edinb).* 2018;120(2):125-137. doi 10.1038/s41437-017-0024-3
2. Ruvinskiy D., Igoshin A. et al. Resequencing the Yaroslavl cattle genomes reveals signatures of selection and a rare haplotype on BTA28 likely to be related to breed phenotypes. *Anim Genet.* 2022;53(5):680-684. doi 10.1111/age.13230-x

doi 10.18699/bgrs2024-5.1-19

Preliminary results of determining the mitochondrial genetic diversity of Ovodov horses in Southern Siberia

Kusliy M.A.[1]*, Malikov D.G.[2], Klementiev A.M.[3], Graphodatsky A.S.[1], Vorobieva N.V.[1], Molodtseva A.S.[1]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Department of Diversity and Evolution of Genomes, Laboratory of Animal Cytogenetics, Novosibirsk, Russia*

[2] *V.S. Sobolev Institute of Geology and Mineralogy, SB RAS, Laboratory of Cenozoic Geology, Paleoclimatology and Mineralogical Indicators of Climate, Novosibirsk, Russia*

[3] *Institute of the Earth's Crust, SB RAS, Irkutsk, Russia*

** kusliy.maria@mcb.nsc.ru*

Key words: ancient DNA; mitochondrial DNA; Ovodov horse; *Equus ovodovi*; extinct species; phylogenetics; genetic diversity

Motivation and Aim: Published genetic data [1–4] on the Ovodov horse made it possible to expand the range of this species, identified by morphometric analysis results, to the northern regions of China and rejuvenate the time of extinction of *Equus ovodovi* to the middle Holocene, while the low degree of genetic diversity of the Ovodov horse Holocene population was identified as the main reason for the extinction. Since the sample of Ovodov horses from Siberia studied at the DNA level (2 samples) was small compared to Ovodov horses from China (21 samples), we expanded it and obtained almost complete mitogenomes (mitochondrial genomes) for 10 Late Pleistocene Ovodov horses from Southern Siberia. The main objectives of the study were to confirm the northern boundary of the range of the Ovodov horse (*Equus ovodovi*), identified on the basis of morphological data (56° north latitude), and to conduct a comparative analysis of the mitochondrial genetic diversity of the Ovodov horse populations of Siberia, China and other populations of wild and domestic species of the genus *Equus*.

Methods and Algorithms: All experiments were conducted in a special laboratory for ancient DNA research, in accordance with the basic authenticity criteria [5]. Ancient DNA isolation from bone samples was carried out according to the protocol described in the article by Vorobiova and co-authors [6]. Ancient DNA fragment libraries were obtained using TruSeq Nano DNA Sample Preparation Kit (Illumina) according to the manufacturer's protocol with a few modifications: sample purification using MinElute PCR Purification Kit (Qiagen), 12 cycles of library amplification. Two rounds of library enrichment were performed based on a previously published method [7]. Paired-end sequencing of the enriched libraries was performed on the MiSeq platform (Illumina) using MiSeq v2 Reagent Kit (300-cycles, 2x150bp). The PALEOMIX BAM Pipeline v1.3.2 [8] was used to trim and collapse reads, align reads against the Ovodov horse mitogenome reference sequences, filter out PCR duplicates, and reassess base quality scores in accordance with the probability of post-mortem damage. Consensus mitogenome sequences were obtained using the bioinformatics software platform Geneious Prime v2024.0.4 with a set threshold of the 60 % highest quality. The phylogenetic tree was constructed using the program MrBayes v.3.2.6 [9] (8 million generations, sampling frequency – 1000, the first 25 % of trees were discarded) based on multiple mitochondrial genome sequence alignment performed in the MAFFT v1.4.0

program [10] and the best-fit partitioning schemes and models of molecular evolution defined in PartitionFinder v2.1.1 [11]. The final phylogenetic tree was visualized using the program FigTree v1.4.4. F_{ST} and nucleotide diversity values were obtained using an integrated software package Arlequin v3.5.2.2 [12].

Results: The graphs of nucleotide misincorporations and the average library fragment sizes for the studied Siberia Ovodov horse samples confirm the ancient origin of the investigated samples. Because the ND1 gene was incompletely assembled in many samples, we excluded it from the multiple alignment, resulting in 98–100 % coverage of the remaining mitogenome sequence for all samples, with an average coverage depth of 47,22 times. The belonging of the samples from Krasnoyarsk to *Equus ovodovi* species confirms the extreme northern border of this species range at 56° north latitude. Based on obtained by us mitogenome sequences of 10 Ovodov horses from Siberia and mitogenome sequences of 123 non-caballine and 122 caballine horses from nucleotide databases (GenBank, the European Nucleotide Archive), we carried out phylogenetic reconstructions. They showed that all the studied Ovodov horses of Siberia are located in clade B of the studied species in the phylogenetic tree, together with all Late Pleistocene and most of the Holocene Ovodov horses from China. The phylogeographic structure is noticeable, since the Ovodov horses of China and Siberia are located in different subclades of clade B, which indicates a certain degree of differentiation of the Ovodov horse populations of these regions. However, the alternation of these clades indicates the interpenetration of the mitochondrial gene pools of these populations during the Late Pleistocene. The F_{ST} analysis confirmed a moderate degree of differentiation between the populations under consideration, which may be due to the fact that the studied horses from China and Southern Siberia lived at opposite extremes of this species range, determined by genetic data. To obtain a more detailed phylogeographic structure of *Equus ovodovi* species, it is necessary to study samples from the central part of the range: the south of Eastern Siberia, Mongolia. We also determined that the value of nucleotide diversity of the studied populations of Ovodov horses in Southern Siberia and China is extremely low and close to that of the populations of wild Iberian horses of the Copper Age, historical quaggas and Ragusano donkeys of the island of Sicily [13, 14], geographically isolated or on the verge of extinction. This most likely indicates that the *Equus ovodovi* species has been endangered on the territory of Siberia and China since at least 40 kya. However, in China the species survived until the mid-Holocene. When, as for the territory of Siberia, we determined only the preservation of representatives of the species a little after the Last Glacial Maximum (end of the Pleistocene). Whether the species survived in Southern Siberia until the middle of the Holocene remains open to question. Perhaps the bone remains of Holocene Ovodov horses from Siberia have not yet been found. Also, such a difference may be due to features in environmental conditions, competition for resources with other species of horses with which their habitats overlapped (the wild horse). In order to find out the genetic diversity of the species under study during its heyday, it is necessary to study the Ovodov horse samples dating back to the period 84–74 kya, the time of its maximum population. However, these samples have not yet been discovered.

Conclusion: Our research showed that the range of the Ovodov horse (*Equus ovodovi*) in Siberia extended north to at least 56° north latitude. Conducted phylogenetic reconstructions and population genetic analysis showed a moderate degree of differentiation between the Ovodov horse populations of Siberia and China, which indicates a certain degree of geographic isolation of the extreme northwestern and

southeastern parts of the range. The mitochondrial genetic diversity we determined in the horse populations of Siberia and China shows that the species was on the verge of extinction since at least 40 kya. Whether the Ovodov horse species survived in Siberia until the mid-Holocene requires new expeditions and genetic research.

Funding: The study is supported by the grant from Russian Science Foundation (No. 23-74-10060, https://rscf.ru/project/23-74-10060/).

References

1. Druzhkova A.S., Makunin A. et al. Complete mitochondrial genome of an extinct Equus (Sussemionus) ovodovi specimen from Denisova cave (Altai, Russia). *Mitochondrial DNA Part B*. 2017;2:79-81. doi 10.1080/23802359.2017.1285209
2. Vilstrup J.T., Seguin-Orlando A. et al. Mitochondrial phylogenomics of modern and ancient equids. *PLoS One*. 2013;8:e55950. doi 10.1371/journal.pone.0055950
3. Yuan J.-X., Hou X.-D., Barlow A. et al. Molecular identification of late and terminal Pleistocene Equus ovodovi from northeastern China. *PLoS One*. 2019;14:e0216883. doi 10.1371/journal.pone.0216883
4. Cai D., Zhu S., Gong M. et al. Radiocarbon and genomic evidence for the survival of Equus Sussemionus until the late Holocene. *Elife*. 2022;11:e73346. doi 10.7554/eLife.73346
5. Willerslev E., Cooper A. Ancient DNA. *Proc R Soc B Biol Sci*. 2005;272:3-16. doi 10.1098/rspb.2004.2813
6. Vorobieva N.V., Makunin A.I. et al. High genetic diversity of ancient horses from the Ukok Plateau. *PLoS One*. 2020;15(11):e0241997. doi 10.1371/journal.pone.0241997
7. Maricic T., Whitten M., Pääbo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*. 2010;5:e14004. doi 10.1371/journal.pone.0014004
8. Schubert M., Jónsson H. et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci*. 2014;111:E5661-E5669. doi 10.1073/pnas.1416991111
9. Ronquist F., Huelsenbeck J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572-1574. doi 10.1093/bioinformatics/btg180
10. Katoh K., Misawa K., Kuma K., Miyata T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30:3059-3066. doi 10.1093/nar/gkf436
11. Lanfear R., Calcott B., Ho S.Y.W., Guindon S. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Mol Biol Evol*. 2012;29:1695-1701. doi 10.1093/molbev/mss020
12. Excoffier L., Lischer H.E.L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10:564-567. doi 10.1111/j.1755-0998.2010.02847.x
13. Cozzi M.C., Valiati P. et al. Mitochondrial DNA genetic diversity in six Italian donkey breeds (Equus asinus). *Mitochondrial DNA Part A*. 2018;29:409-418. doi 10.1080/24701394.2017.1292505
14. Lorenzen E.D., Arctander P., Siegismund H.R. High variation and very low differentiation in wide ranging plains zebra (Equus quagga): insights from mtDNA and microsatellites. *Mol Ecol*. 2008;17:2812-2824. doi 10.1111/j.1365-294X.2008.03781.x

doi 10.18699/bgrs2024-5.1-20

Molecular genetic response to oxidative stress differs in hypothalamus of hypertensive ISIAH and normotensive WAG rats

Makovka Yu.V.[1, 2]*, Oshchepkov D.Yu.[1], Fedoseeva L.A.[1], Markel A.L.[1, 2], Redina O.E.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *makovkayv@bionet.nsc.ru*

Key words: hypothalamus; RNA-Seq; gene expression; response to oxidative stress; restraint stress; ISIAH rat strain

Motivation and Aim: It is known that an important link in the pathogenesis of hypertension is oxidative stress, which occurs under the influence of various environmental or endogenous factors, including psycho-emotional stress. The response to psycho-emotional stress may depend on the genotype and physiological state of the organism as a whole. However, the features of the molecular genetic mechanisms of the response to oxidative stress depending on the genotype have not been fully studied. The purpose of this work was to establish the inter-strain hypothalamic differences in the expression of genes associated with the response to oxidative stress in hypertensive ISIAH and normotensive WAG rats after exposure to a single two-hour restraint (emotional) stress.

Methods and Algorithms: Male (3 months old) hypertensive ISIAH rats and normotensive WAG rats were used in the study. The animals were divided into 4 groups: ISIAH_control, WAG_control, ISIAH_stress, WAG_stress (n=7 in each group). The basal systolic blood pressure was measured in all animals by the tail-cuff method using ether anesthesia [1]. After 7 days, the rats from the experimental groups were subjected to restraint stress, for which the rats were placed in a tight wire-mesh cage for 2 hours. Upon completion of the stress procedure, blood pressure was measured in these rats without anesthesia, after which immediate decapitation and isolation of the hypothalamus were performed. RNA-Seq analysis of hypothalamic samples was carried out on the DNBSEQ platform (DNBSEQ Technology) at BGI Hongkong Tech Solution NGS Lab. Pair-end sequencing of cDNA libraries was performed with a read length of 150 base pairs and sequencing depth of more than 30 million uniquely mapped reads. Bioinformatics analysis included reads mapping to the reference rat genome mRatBN7.2/rn7 using Star package [2], calculation of significant surrogate variables using the SVA package [3] to account for undesirable systematic variation in the sample preparation process, which were further included as factors in the differential expression analysis performed in DESeq2 [4] with the threshold significance set to FDR adjusted p-value<5 %. For functional annotation of differentially expressed genes (DEGs), the Rat Genome Database and DAVID (The Database for Annotation, Visualization and Integrated Discovery) databases were used. The gene set enrichment analysis for promoter binding sites for transcription factors was carried out using the web-based tool Enrichr [5]. Validation of differential expression of target genes was carried out using

real-time PCR (q-PCR). The study protocol was approved by the Bioethical Council of the Federal Research Center The Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia), protocol No. 115 of December 20, 2021.

Results: Functional annotation of DEGs revealed two groups of genes representing (1) a response to oxidative stress common to the two strains (59 DEGs) and (2) oxidative stress response genes specific to ISIAH rats (111 DEGs). DEGs from the common group change the transcription level unidirectionally in rats of both strains. However, in the hypothalamus of normotensive WAG rats, the response to oxidative stress is much less pronounced. Among the common DEGs, the genes Nos1, Ppargc1a, Abcc1, Srxn1, Cryab, Hspb1 and Fosl1 changed the transcription level most significantly. The gene set enrichment analysis for promoter binding sites for transcription factors revealed two terms associated with binding sites for the transcription factor CREB1 and the glucocorticoid receptor (NR3C1) in the hypothalamus of both rat strains, suggesting that these transcription factors may play a key role in regulation of the response to oxidative stress regardless of the genotype/strain differences and hypertensive state of the rats. A specific response to oxidative stress in hypertensive ISIAH rats is associated with induction of Fos gene expression, which was confirmed by q-PCR.

Conclusion: This study demonstrates both common and strain specific features in the molecular genetic mechanisms of the hypothalamic response to oxidative stress in hypertensive ISIAH rats being a model of stress-sensitive form of hypertension and normotensive WAG rats, complementing existing understanding of this process depending on the genotype and physiological state of the body.

Funding: The study is supported by the Russian Science Foundation (grant No. 22-14-00082).

References

1. Markel A.L., Redina O.E., Gilinsky M.A. et al. Neuroendocrine profiling in inherited stress-induced arterial hypertension rat strain with stress-sensitive arterial hypertension. *J Endocrinol.* 2007;195(3):439-450. doi 10.1677/JOE-07-0254
2. Dobin A., Davis C.A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21. doi 10.1093/bioinformatics/bts635
3. Leek J.T., Johnson W.E. et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882-883. doi 10.1093/bioinformatics/bts03
4. Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi 10.1186/s13059-014-0550-8
5. Xie Z., Bailey A., Kuleshov M.V. et al. Gene set knowledge discovery with enrichr. *Curr Protoc*. 2021;1(3):e90. doi 10.1002/cpz1.90

doi 10.18699/bgrs2024-5.1-21

Testing the 'parasite-mediated domestication' hypothesis: a comparative approach to the wild boar and domestic pig as model species

Oleinic R.[1], Posedi J.[2], Beck R.[3], Šprem N.[4], Škorput D.[5], Pokorny B.[6, 7], Škorjanc D.[1], Prevolnik Povše M.[1], Skok J.[1]*

[1] *Department of Animal Science, Faculty of Agriculture and Life Sciences, University of Maribor, Maribor, Slovenia*

[2] *Unit for Parasitology, Institute of Microbiology and Parasitology, Veterinary Faculty, University of Ljubljana, Ljubljana, Slovenia*

[3] *Laboratory for Parasitology, Croatian Veterinary Institute, Zagreb, Croatia*

[4] *Department of Fisheries, Apiculture, Wildlife Management and Special Zoology, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia*

[5] *Division of Animal Science, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia*

[6] *Faculty of Environmental Protection, Velenje, Slovenia*

[7] *Slovenian Forestry Institute, Ljubljana, Slovenia*

* *janko.skok@um.si*

Key words: domestication; domestication syndrome; endoparasites; helminths; protozoa; *Sus scrofa*

Motivation and Aim: Many of the mechanisms behind the domestication/domestication syndrome have already been well explained. Starting with the pioneering studies on domestication by Dmitry Belyaev, who proposed that the domestication syndrome is genetically linked to genes associated with tameness. His experiment on domestication of silver foxes, which is among the most influential work in this field, showed that selection for tameness (impaired stress response, changes in the 'hypothalamic-pituitary-adrenal system' – HPA axis) leads to significant destabilisation of regulatory systems controlling morphological and behavioural development, resulting in changes that are otherwise characteristic of domestication syndrome [1]. Much later, some other findings were added. In the thyroid rhythm hypothesis, Crockford [2] proposed that domestication is also driven by genetically controlled changes in the activity rhythm of the thyroid gland, which have a crucial effect on heterochronic changes and thus play an important role in the domestication syndrome (e. g. paedomorphism). Furthermore, Wilkins et al. [3] proposed that the main phenotypic components of domestication syndrome are neural crest cells (NCC) derivatives, i. e. a result of a developmental reduction in NCC input for the affected phenotypic traits. Recently, the parasite-mediated domestication hypothesis (PMD) has been proposed [4]. Indeed, parasites literally affect all the major mechanisms that otherwise underlie the domestication syndrome (HPA, thyroid, NCC – e. g. neuroendocrine regulation, hormonal modulation, changes in the host miRNA profile, etc.). In addition, the characteristics of the domestication syndrome could be genetically linked to genes related to resistance/tolerance to parasites. Therefore, the PMD assumes an important role of endoparasites in the process of domestication, especially in the initial phase (proto-domestication). It predicts that the frequency of domestication syndrome traits in the wild population increases with decreasing genetic resistance to parasites and/or with

increasing parasite load. PMD can be tested in different ways, either experimentally (experimentally parasitised population) or comparatively (existing wild and domestic populations). Although comparative studies can be problematic for a variety of reasons, they are relatively easy to conduct and are therefore suitable for an initial test of hypothesised PMD baselines. It can be assumed that, under comparable conditions, the parasite load in the domestic population will be higher than in the wild population. Therefore, we tested the PMD here with a systematic comparative approach by analysing the parasite load in wild boar and free-ranging domestic pigs from a comparable environment.

Methods: Wild boar populations and free-ranging (grazing) domestic pig populations, one each from Slovenia (SLO) and Croatia (CRO), were included in the study. In both countries, the wild boar and domestic pig populations were located about 15 km apart, so they had similar chances of being exposed to the local parasitofauna. Domestic pigs were not threated against parasites. Sampling of fresh faeces (from the rectum of hunted wild boar and immediately after defecation in domestic pig) was carried out between 3 November and 11 December 2023. A total of 59 individual faecal samples (SLO: 12 wild boar, 20 domestic pig; CRO: 14 wild boar, 13 domestic pig) were examined for endoparasites. The endoparasite load was analysed using faecal flotation and sedimentation diagnostics.

Results: A total of five different parasite taxa were found in the CRO samples. In terms of parasite load, there were no differences between the wild and domestic stratum, which were each infected with three different parasite taxa (wild boar: *Eimeria* sp., strongyle-type eggs, and *Strongyloides* spp.; domestic pig: *Cystoisospora suis*, strongyle-type eggs, and *Trichuris* sp.). There was no statistical difference between the strata in the strongyle-type eggs found in wild boar and domestic pig. A total of four different parasite taxa were found in SLO samples. Only two of them were found in wild boar (*Eimeria* sp., and *Oesophagostomum* sp.), but all four were found in domestic pig (*Eimeria* sp., *Balantidium coli*, *Oesophagostomum* sp., *Trichuris* sp.). For two matching parasites the load of *Oesophagostomum* sp. was significantly higher in domestic pig than in wild boar.

Conclusion: According to our preliminary results, there are indications in favour of PMD. However, we cannot draw a firm conclusion as there are many aspects that can mislead the interpretation. The first aspect concerns the possible artificial (either intentional or spontaneous) selection of resistant/tolerant animals aimed at increasing resistance [5] and possibly leading to a deceptively high resistance/tolerance to parasites in the domestic population, which was certainly not the case in our study. The second aspect, which may lead to the opposite, concerns relaxed selection, where the source of selection that was previously important for the maintenance of a particular trait is weakened or even eliminated [6]. Indeed, selection pressure from parasites may have been reduced during the domestication process, which, together with "stringent" artificial selection for desirable traits, could influence (jeopardise) the evolution of genetic resistance in (parasite-naïve) domestic animals [7, 8]. In pigs, however, it has been shown that domestication does not appear to act as a bottleneck limiting the diversity of parasite resistance genes [9]. Testing PMD based on parasite load/resistance/tolerance in existing populations of wild or domestic animals and simply comparing these populations is therefore interpretatively quite complex, although relatively easy to perform. Comparative studies, such as the present one, should be supported by a more focussed methodology. Either to examine the frequency of domestication syndrome traits in the wild population in relation to their parasite

resistance/load, or to examine the parasite resistance/load of wild animals showing signs of domestication syndrome (e. g. tameness) in comparison to completely wild animals of the same population. The other possibility would be an experimental approach, i. e. experimental proto-domestication, like the Belyaev fox experiment [1]. However, instead of selecting the animals for tameness, an experimental population of the wild counterpart of the domestic animals would be experimentally exposed to parasites, selected against parasite resistance, and analysed over generations for the frequency of domestication syndrome traits in the population. Besides, it must also be considered that the influence of parasites on the domestication process may not simply be generalised, as it is also possible that only certain parasites play a mediating role in the domestication process.

Funding: Our special thanks go to the domestic pig breeders Bojan Lešnik (Slovenia) and Željko Horvat (Croatia), who made it possible for us to obtain samples from domestic pigs, and to Luka and Boštjan Žunkovič, who enabled access to samples from wild boar in Slovenia. The study was partially supported by the Slovenian Research and Innovation Agency (research programmes No. P1-0164, P4-0092, P4-0107), Ministry of Agriculture, Forestry and Food of Republic of Slovenia (V4-2223), and Croatian Science Foundation HRZZ (IP-2022-10-7502).

References

1. Belyaev D.K. Destabilizing selection as a factor in domestication. *J Heredity*. 1979;70(5):301-308. doi 10.1093/oxfordjournals.jhered.a109263
2. Crockford S.J. Animal domestication and vertebratespeciation: a paradigm for the Origin of Species. Doctoral Dissertation. British Columbia, Canada: University of Victoria, 2004
3. Wilkins A.S., Wrangham R.W., Fitch W.T. The "domestication syndrome" in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*. 2014;197(3):795-808. doi 10.1534/genetics.114.165423
4. Skok J. The Parasite-Mediated Domestication Hypothesis. *Agric Scientia*. 2023;20(1):1-7. doi 10.18690/agricsci.20.1.1
5. McManus C., do Prado Paim T., de Melo C.B., Brasil B.S., Paiva S.R. Selection methods for resistance to and tolerance of helminths in livestock. *Parasite*. 2014;21:56. doi 10.1051/parasite/2014055
6. Lahti D.C., Johnson N.A., Ajie B.C., Otto S.P., Hendry A.P., Blumstein D.T., Coss R.G., Donohue K., Foster S.A. Relaxed selection in the wild. *Trends Ecol Evol*. 2009;24(9):487-496. doi 10.1016/j.tree.2009.03.010
7. Eizaguirre C., Lenz T.L., Kalbe M., Milinski M. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nat Commun*. 2012;3(1):621. doi 10.1038/ncomms1632
8. Smallbone W., Ellison A., Poulton S., van Oosterhout C., Cable J. Depletion of MHC supertype during domestication can compromise immunocompetence. *Mol Ecol*. 2021;30(3):736-746. doi 10.1111/mec.15763
9. Moutou K.A., Koutsogiannouli E.A., Stamatis C., Billinis C., Kalbe C., Scandura M., Mamuris Z. Domestication does not narrow MHC diversity in Sus scrofa. *Immunogenetics*. 2013;65:195-209. doi 10.1007/s00251-012-0671-8

doi 10.18699/bgrs2024-5.1-22

Relationships between morphological, physiological and behavioral traits associated with the age-dependent development of stress-sensitive arterial hypertension in the ISIAH rats

Oshchepkov D.Yu.[1, 3]*, Makovka Yu.V.[1, 2], Ponomarenko M.P.[1], Redina O.E.[1], Markel A.L.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *diman@bionet.nsc.ru*

Key words: ISIAH rat strain; age-related difference; hypertension; principle component analysis; stress reactivity

Motivation and Aim: Hypertension is one of the most significant risk factors for many cardiovascular diseases. At different stages of arterial hypertension development, various pathophysiological processes can play a key role in the manifestation of the hypertensive phenotype and comorbid conditions. Accordingly, it is thought that when diagnosing and choosing a treatment strategy for arterial hypertension, it is necessary to take into account age, stage of disease development, comorbidities and the influence of emotional and psychosocial factors. However, this approach to treatment selection is hampered by incomplete knowledge of the details of the age-related associations between numerous features that may contribute to the manifestation of the hypertensive phenotype. Therefore, we carried out a study to analyze the age dependent changes in relationships between morphological, physiological and behavioral traits in two groups of hybrid F2(ISIAHxWAG) male rats of different ages.

Methods and Algorithms: The study used two groups of F2(ISIAHxWAG) hybrid males aged 3 (n=103) and 6 months (n = 126), obtained by crossing hypertensive ISIAH rats (simulating stress-sensitive form of arterial hypertension) and normotensive WAG rats. Using a group of hybrid rats allows us to obtain the full range of variability between normal and hypertensive phenotypes. The data set we obtained for these two groups of rats includes data on morphological, physiological and behavioral traits, many of which show interstrain differences and represent the key features in the development of the hypertensive phenotype. Using the principal component method, the relationships between 21 traits were analyzed. The results of previous studies have shown that behavioral traits of rats in the open field test and morphometric traits contributed to different components [1]. Taking this into account, in the present work, the analysis of behavioral and morphometric traits was carried out separately.

Results: The most significant results were obtained when analyzing the set of behavioral and physiological characteristics for rats of both ages taken together. It has been shown, that the development of stress-sensitive hypertension in ISIAH rats is accompanied not only by an age-related (FDR<5 %) persistent increase in basal blood pressure, but also by a decrease in the response to stress and an increase in anxiety: the age indicator was positively and significantly correlated only with second principal component (PC2). The

main contribution to PC2, accounting for 15.4 % of the total variance, was made by the characteristics of hypertensive status, with a positive contribution from basal blood pressure, which suggests that the development of hypertensive status is the main age-related parameter by which the groups of F2(ISIAHxWAG) males differ. The negative contribution of an increase in blood pressure during stress to PC2 corresponds to a decrease in glucocorticoid stress response with age, which is consistent with data obtained in earlier studies. A number of behavioral characteristics also correlated with PC2: a negative contribution was noted for locomotor activity, and a positive contribution was found for parameters such as displaced activity (grooming) and the latency period. That is, in this case we are dealing with behavioral parameters that characterize the conflict of several motivations: anxiety, fear and exploration. Consequently, the second component, associated with the age of rats, can be characterized as a factor of behavioral caution (indecisiveness) and anxiety, which increases with age and the development of sustained hypertension [2]. Analysis performed for the two age groups separately showed that the plasma corticosterone concentration at rest and its increase during short-term restraint stress in a group of young rats did not have a straightforward relationship with the other analyzed traits.
However, in contrast to young rats, such associations were found in older animals, suggesting that glucocorticoid adrenal function may modulate a range of behavioral traits associated with emotionality in later life.
Conclusion: The study revealed age-dependent relationships between the key features that determine the manifestation of hypertension in ISIAH rats. Our results may be useful for developing the specific protocols of the treatment strategies for stress-sensitive hypertension, taking into account the age of patients.
Funding: The study is supported by the Russian Science Foundation (grant No. 22-14-00082).

References

1. Efimov V.M., Kovaleva V.Y., Markel A.L. A new approach to the study of genetic variability of complex characters. *Heredity*. 2005;94(1):101-107. doi 10.1038/sj.hdy.6800580
2. Oshchepkov D.Y., Makovka Y.V., Ponomarenko M.P., Redina O.E., Markel A.L. Age-dependent changes in the relationships between traits associated with the pathogenesis of stress-sensitive hypertension in ISIAH rats. *Int J Mol Sci*. 2023;24(13):10984. doi 10.3390/ijms241310984

doi 10.18699/bgrs2024-5.1-23

Expression of glutamatergic system genes in rats with stereotypies and audiogenic epilepsy (PM strain)

Plekanchuk V.S.*, Ryazanova M.A.

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

** lada9604@mail.ru*

Key words: glutamatergic system; gene expression; PM rat strain; audiogenic epilepsy; catatonia

Motivation and Aim: PM strain (an abbreviation for "pendulum-like movements") was obtained by selection for increased pendulum-like movements as a model of stereotypies in catatonic syndrome. In addition it has been noticed that PM rats also have a predisposition to seizures caused by audiogenic stimuli [1]. Considering the data on the involvement of the glutamatergic system in the etiopathogenesis of eplepsy and catatonia, the aim of the work was set: to study the expression of the glutamatergic system genes at the transcriptional level of the PM rats brain in comparison with Wistar controls.

Methods and Algorithms: Male rats of the PM and Wistar strains at the age of 6 months were used for the study. Using RT-qPCR *Grin1*, *Grin2A*, *Grin2B*, *Gria1*, *Grm2*, *Grm3*, *Grm5*, *Slc1a3*, *Slc1a2*, *Gad1*, *Gls*, *Slc17a6* mRNA level in brain structures (hippocampus, prefrontal cortex, striatum,) of PM and Wistar rats was estimated.

Results: The study revealed a high level of *Grm2* mRNA in hippocampus and prefrontal cortex of PM rats. This gene encodes a metabotropic autoreceptor, which, through a negative feedback mechanism, suppresses the release of glutamate into the synaptic cleft. In addition, in PM rats, an increase in the mRNA of another metabotropic receptor gene, *Grm5*, was shown in the prefrontal cortex. This postsynaptic receptor enhances glutamate flux through the NMDA receptors. There is evidence in the literature about the contribution of metabotropic receptors to the development of stress-related disorders [2]. Compared to the control, the level of mRNA of the *Slc1A3* gene of the EAAT1, responsible for the reuptake of glutamate, is higher in the hippocampus of PM rats, which may influence the amount of glutamate in the synaptic cleft. Data were obtained on a significant decrease in the hippocampus of PM rats in the mRNA of the *Gad1* gene, encoding the enzyme for the synthesis of GABA from glutamate. It is the imbalance of excitatory and inhibitory neurotransmitters that is believed to underlie psychopathological or neurological symptoms, including epilepsy [3].

Conclusion: The identified changes in the glutamate system genes expression in PM rats brain structures indicate changes in the neurotransmission of glutamate and can affect the level of excitation of these brain structures and electroconvulsive activity. The data obtained may help to identify the mechanisms of development of stereotypical pendulum-like movements and the susceptibility to audiogenic seizures of PM rats.

Funding: The study is supported by budget project No. FWNR-2022-0019.

References

1. Kolpakov V.G. Catatonia in animals: genetics, neurophysiology, neurochemistry. Novosibirsk: Nauka, 1990 (in Russian)
2. Dogra S., Conn P.J. Targeting metabotropic glutamate receptors for the treatment of depression and other stress-related disorders. *Neuropharmacology*. 2021;196:108687. doi 10.1016/j.neuropharm.2021.108687
3. Akyuz E., Polat A.K., Eroglu E. et al. Revisiting the role of neurotransmitters in epilepsy: An updated review. *Life Sci.* 2021;265:118826. doi 10.1016/j.lfs.2020.118826

doi 10.18699/bgrs2024-5.1-24

Transcriptomic response of genes related to the glutamate- and GABA-ergic systems in the hypothalamus of ISIAH and WAG rats exposed to acute restraint stress

Plekanchuk V.S.[1]*, Ryazanova M.A.[1], Oshchepkov D.Yu.[1, 2], Redina O.E.[1], Markel A.L.[1, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center ICG SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *lada9604@mail.ru*

Key words: glutamatergic system; GABA-system; hypothalamus; stress; gene expression; arterial hypertension; RNA-Seq; ISIAH rat strain

Motivation and Aim: The hypothalamus is the main structure of the brain that implements the body's response to stress by activating the hypothalamic-pituitary-adrenal axis (HPA) and other systems. Glutamate and GABA are the primary neurotransmitters that provide the necessary level of excitation of the hypothalamic neuroendocrine cells. An imbalance of these neurotransmitters can lead to significant changes in HPA axis activity [1]. The aim of the work was to compare changes in the transcriptional activity of genes related to the glutamate- and GABA-ergic systems of the hypothalamus in response to acute restriction stress in ISIAH rats modeling a stress-sensitive form of arterial hypertension and in normotensive WAG rats.

Methods and Algorithms: Male rats of the hypertensive line ISIAH/Icgn and the normotensive line WAG/GSto-Icgn at the age of 3 months were used for the study. Transcriptomic sequencing (RNA-Seq) of the hypothalamus was performed on 4 groups of rats: 1) ISIAH_control; 2) WAG_control; 3) ISIAH_stress; 4) WAG_stress (7 rats in each group). Two groups of rats (ISIAH and WAG) were subjected to restriction stress by placing the rats in a tight wire-mesh cage for 2 hours. After completing the stress procedure, the rats were immediately decapitated, the hypothalamus was isolated, and the samples were homogenized in the ExtractRNA reagent (Evrogen, Russia). Paired-end sequencing of cDNA libraries was performed on the DNBSEQ platform (DNB-SEQ technology) with a read length of 150 bp and a sequencing depth of more than 30 million uniquely mapped reads. All samples were analyzed as biological replicates. The study protocol was approved by the Bioethical Council of the federal research center Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia), protocol No. 115 of December 20, 2021. The quality of the sequencing data was assessed using FastQC software (version 0.11.5). Sequencing data were mapped to the rat reference genome mRatBN7.2/rn7 (Wellcome Sanger Institute, rn7 assembly, November 2020) using the STAR software package, version 2.7.10a. Differential gene expression (DEG) assessment was performed with DESeq2 v1.30.1 using surrogate variable analysis (SVA) to account for unwanted variations in the data caused by possible random biases during sample preparation. The significance threshold for determining DEGs was chosen taking into account correction for multiple comparisons and corresponded to adjusted p-

value < 5%. The results show only DEGs that changed their transcription level under stress by more than 1.5-fold (Log2 fold change ≥|0.585|).

Results: Unidirectional downregulation of the expression of the *Gad2, Grik3*, and *Gabrb3* genes in the hypothalamus of both strains of rats indicates a nonspecific (rat strain-independent) change in their expression in response to restraint stress. The *Gad2* gene encodes the enzyme for the conversion of glutamate to GABA [2], and a decrease in the amount of its mRNA can lead to a change in the balance of excitation and inhibition processes. The *Grik3* gene encodes a subunit of the kainate glutamate ionotropic receptor GluK3. Kainate receptors are involved in postsynaptic signal transduction and control of neuronal excitability, and they presynaptically modulate the release of GABA and glutamate [3]. A decrease in *Grik3* gene expression in response to acute stress may lead to changes in GABAergic and glutamatergic transmission in hypothalamus of both ISIAH and WAG rat strains. The *Gabrb2* and *Gabrb3* genes encode metabotropic GABA-B receptors. The amount of *Gabrb3* gene mRNA decreased in rats of both strains under acute stress, while a decrease in the amount of *Gabrb2* mRNA occurred only in the hypothalamus of ISIAH rats. Exposure to stress has shown that the greatest inter-strain differences are observed in changes in the transcriptional activity of the GABA-A receptor gene subunits. The *Gabrq* and *Gabrd* genes encode the δ and θ subunits respectively. This study shows a decrease in *Gabrq* mRNA and an increase in *Gabrd* mRNA after stress in the hypothalamus of WAG, but not ISIAH, rats. The θ-subunit has not been sufficiently studied, while the δ-subunit has been shown to play an important role in the regulation of hypothalamic neurons that synthesize corticotropin-releasing hormone, and knockout of the gene encoding this subunit in these neurons reduces the corticosteroid response to stress, probably due to loss of the disinhibitory effect of GABA following acute stress [4]. Our work also shows a different response to acute stress at the level of transcriptional activity of the metabotropic glutamate receptor gene *Grm5*: a decrease in mRNA in the hypothalamus of ISIAH rats, but no changes in WAG. The contribution of this receptor to synaptic plasticity, the development of long-term potentiation, and the regulation of cell excitability is described in the literature [5].

Conclusion: Short-term (2 hours) restraint stress causes changes in the expression of a significant number of genes related to the functioning of the glutamate and GABAergic systems of the hypothalamus. Downregulation of the expression of the *Gad2, Grik3* and *Gabrb3* genes was shown in rats of both strains and, accordingly, does not depend on the genotype. The strain-specific response to stress in the hypothalamus of hypertensive ISIAH rats is associated with a decrease in the transcription of the *Grm5* and *Gabrb2* genes, and in the hypothalamus of normotensive WAG rats – with a change in the level of transcription of the *Gabrq* and *Gabrd* genes. The identified interstrain differences may modulate the activity of synaptic transmission in the hypothalamus and contribute to different stress reactivity in normotensive WAG and ISIAH rats, which are an animal model a stress-sensitive form of arterial hypertension.

Funding: The study is supported by the Russian Science Foundation (grant No. 22-14-00082).

Acknowledgments: The bioinformatics analysis was conducted using the computational resources of HPC facilities at collaborative center «Bioinformatics» ICG SB RAS.

References

1. Lu J., Li Q., Xu D., Liao Y., Wang H. Programming of a Developmental Imbalance in Hypothalamic Glutamatergic/GABAergic Afferents Mediates Low Basal Activity of the Hypothalamic-Pituitary-Adrenal Axis Induced by Prenatal Dexamethasone Exposure in Male Offspring Rats. *Toxicology Letters*. 2020;331:33-41. doi 10.1016/j.toxlet.2020.05.022
2. Kim K., Yoon H. Gamma-Aminobutyric Acid Signaling in Damage Response, Metabolism, and Disease. *Int J Mol Sci*. 2023;24(5):4584. doi 10.3390/ijms24054584
3. Negrete-Díaz J.V., Falcón-Moya R., Rodríguez-Moreno A. Kainate Receptors: From Synaptic Activity to Disease. *FEBS J*. 2022;289(17):5074-5088. doi 10.1111/febs.16081
4. Lee V., Sarkar J., Maguire J. Loss of Gabrd in CRH Neurons Blunts the Corticosterone Response to Stress and Diminishes Stress-Related Behaviors. *Psychoneuroendocrinology*. 2014;41:75-88. doi 10.1016/j.psyneuen.2013.12.011
5. Bodzęta A., Scheefhals N., MacGillavry H.D. Membrane Trafficking and Positioning of mGluRs at Presynaptic and Postsynaptic Sites of Excitatory Synapses. *Neuropharmacology*. 2021;200:108799. doi 10.1016/j.neuropharm.2021.108799

doi 10.18699/bgrs2024-5.1-25

Distinctive hypothalamic gene expression profiles in hypertensive ISIAH and normotensive WAG rats exposed to a single short-term restraint stress

Redina O.[1]*, Oshchepkov D.[1], Makovka Yu.[1, 2], Fedoseeva L.[1], Seryapina A.[1], Markel A.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

Key words: hypothalamus; short-term restraint stress; gene expression; hypertension; RNA-Seq; ISIAH rat strain

Motivation and Aim: An increase in blood pressure is one of the adaptive manifestations of the stress response, which can represent a significant health risk factor in modern human lifestyles. There is no doubt that there is a relationship between psychoemotional stress and hypertension. However, the study of the genetic basis of the increased blood pressure response to psychoemotional stress remains incompletely studied to date. The study was aimed to identify the strain-specific neural networks providing differences in hypothalamic response to stress in hypertensive ISIAH and normotensive WAG rats.
Methods and Algorithms: The experiment was carried out on three-month-old male hypertensive ISIAH/Icgn and normotensive WAG/GSto-Icgn rats. Transcriptome sequencing (RNA-Seq) of the hypothalamus was performed on 4 groups of rats: (1) ISIAH_control; (2) WAG_control; (3) ISIAH_stress; (4) WAG_stress. There were 7 rats in each group. In all rats, basal systolic blood pressure was measured using the tail-cuff method under light ether anesthesia to prevent emotional stress during the measurement. Seven days after measuring basal blood pressure, two groups of rats (ISIAH and WAG) were subjected to restraint stress by placing the rat in a tight wire mesh cage for 2 hours. After completion of the stress procedure, blood pressure was measured by a tail cuff method without anesthesia, and rats were immediately decapitated. The hypothalamus was isolated and homogenized in the ExtractRNA reagent (Evrogen, Russia). Plasma corticosterone concentrations were measured using the immunoassay (Corticosterone ELISA Kit, Abcam). Pair-end sequencing of cDNA libraries was performed on the DNBSEQ platform (DNB-SEQ Technology) with a read length of 150 base pairs and sequencing depth of more than 30 million uniquely mapped reads. All samples were analyzed as biological replicates. The study protocol was approved by the Bioethical Council of the federal research center Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia), protocol No. 115 of December 20, 2021. FastQC software (version 0.11.5) was used to assess the quality of sequencing data. Sequencing data were mapped to the rat reference genome mRatBN7.2/rn7 (rn7 assembly Wellcome Sanger Institute Nov, 2020) using the STAR software package, version 2.7.10a. Differential gene expression (DEG) assessment was performed with DESeq2 v1.30.1 using surrogate variable analysis (SVA) to account for unwanted variations in the data caused by possible random biases during sample preparation. The significance threshold for determining DEGs was chosen taking into account correction for multiple comparisons and corresponded to adjusted p-value < 5 %.

Results: Groups of DEGs that describe the general response to stress, independent of the genotype of rats, as well as strain-specific responses of the hypothalamus to stress, were identified in ISIAH_stress/ISIAH_control and WAG_stress/WAG_control comparisons. Functional annotation of DEGs made it possible to identify the biological processes that change most significantly in each rat strain. It is assumed that different degrees of activation of oxidative phosphorylation and manifestations of oxidative stress may be significant events in the formation of strain-specific features of the hypothalamic response to stress. It is suggested that alternative changes in the expression of the *Npas4* gene (neuronal PAS domain protein 4), which is downregulated in the hypothalamus of control WAG rats and induced in the hypothalamus of hypertensive ISIAH rats, is a key phenomenon for understanding inter-strain differences in the hypothalamic stress response in these rat strains. The strain-specific response to stress in the hypothalamus of ISIAH rats is characterized by downregulation of most DEGs. At the same time, induction of transcription of the *Fos* and *Jun* genes may play a decisive role in the activation of neuronal functions in these rats. Taking into account our previous data that the induction of *Fos* gene transcription occurs in concert with an increase in blood pressure in ISIAH rats [1], we can assume the existence of an association between the functioning features of the hypothalamic gene networks we identified and a sharp increase in blood pressure in ISIAH rats exposed to stress.
Conclusion: Distinctive differential expression of a large number of genes is observed in hypothalamic profiles of hypertensive ISIAH and normotensive WAG rats exposed to single short-term restraint stress. The data obtained may be potentially useful in the selection of molecular targets for the development of pharmacological approaches to the correction of stress-induced pathologies associated with neuronal excitability, taking into account the hypertensive status of patients.
Funding: The study is supported by the Russian Science Foundation (grant No. 22-14-00082).

References

1. Makovka Y.V., Fedoseeva L.A., Oshchepkov D.Y., Markel A.L., Redina O.E. Restraint stress-induced expression of *Fos* and several related genes in the hypothalamus of hypertensive ISIAH rats. *Mol Biol.* 2024;58(1):62-70. doi 10.1134/S0026893324010072

doi 10.18699/bgrs2024-5.1-26

Phylogenetic reconstructions performed for ancient and modern individuals belonging to Canidae family from the territory of Siberia and Central Asia

Samarina S.[1, 2]*, Molodtseva A.[2], Kusliy M.[2], Serdyukova N.[2], Perelman P.[2], Beklemisheva V.[2], Malikov D.[2, 3], Klementiev A.[4], Lurie P.[5]

[1] *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

[2] *Laboratory of Animal Cytogenetics, Department of Diversity and Evolution of Genomes, Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[3] *Laboratory of paleoclimatology, Cenozoic geology and mineral indicators of climate, Sobolev Institute of Geology and Mineralogy, SB RAS, Novosibirsk, Russia*

[4] *Laboratory of Cenozoic, Institute of the Earth Crust, SB RAS, Novosibirsk, Russia*

[5] *Oriental Department, The State Hermitage Museum, St. Petersburg, Russia*

* *s.samarina@g.nsu.ru*

Key words: Canidae; ancient; *Vulpes vulpes*; *Canis lupus*; *Cuon alpinus*; *Vulpes corsac*; phylogenetic tree; phylogenetic reconstruction; haplogroup network; mtDNA; mitochondrial genome; fox; wolf; dhole; corsac

Motivation and Aim: Our work included ancient and recent animals of three extant species from different sites of the Siberia region and an ancient sample of *Cuon alpinus*, extinct species of Siberia. Except for the wolf (*Canis lupus*) and the dog (*Canis lupus familiaris*) species, the data concerning mtDNA sequences of Canidae species in open databases is poor and often restricted by Eastern Asia or Europe regions. Our goal is to add some data about mtDNA sequences of Canidae species for future comprehensive phylogenetic analyses. Existing phylogenetic analyses for *Vulpes vulpes, Vulpes corsac* and *Cuon alpinus* are based on some short regions of mtDNA sequence and on the restricted part of an area. Moreover, full ancient mitogenomes are received only for *Canis lupus*, the history of other species' evolution is to be revealed in future. Hence, the aim of the study is to define the phylogenetic connections among ancient and modern representatives of canids from the regions of Siberia and Central Asia.

Methods and Algorithms: In our work ancient samples were collected by paleontologists from different regions of Siberia, such as the Omsk Oblast, the Republic of Khakassia, the Altai Republic, the Irkutsk Oblast, the Krasnoyarsk Krai. One sample was collected from the location in Tajikistan. Modern samples were collected from the Novosibirsk Oblast, the Russian Far East and the Sakha Republic. Ancient DNA was isolated according to Yang's protocol [1] with modifications [2], the libraries for high-throughput sequencing were constructed using TruSeq® Nano DNA Sample Preparation Kit with index sequences from Truseq DNA HT Library Preparation Kit or Truseq Nano DNA LT Sample Preparation Kit (Illumina). According to the protocol of Maricic et al. [3] with modifications [4] enrichment with target fragments in two rounds was carried out. Modern libraries were enriched in one round. Consensus sequences were obtainedby alignments against reference sequences. Then models of evolution for every subset in the genome were determined, models of damage were contructed by MapDamage v2.2.0 package [5]. The phylogenetic tree for Canidae was built using the MrBayes program

[6]; the haplogroup median-joining network for ancient and modern fox species was constructed in the Network program (Fluxus Technology Ltd).

Results: We have constructed 34 DNA libraries from ancient and modern isolated DNA samples. Only few ancient samples demonstrated enough mitogenome coverage to use in the phylogenetic reconstruction. However, all 9 modern samples achieved enough coverage (more than 94 %). The haplogroup network was obtained showing some new Siberian haplotypes and some known from available literature haplotypes of the cytochrome B gene. For *Vulpes corsac* there was only one available mitogenome in GenBank database, so we obtained another three haplotypes for this species. An ancient sample took the special position in both reconstructions which was relevant to its geographical location. We gained 68 % of mitogenome coverage for *Cuon alpinus* sample and confirmed its species affiliation. We look forward to increase this value to get the first complete or nearly complete mitogenome of ancient *Cuon alpinus*.

Conclusion: The phylogenetic position for a number of ancient and modern individuals of Canidae family was determined. Furthermore, the group obtained the decent mitogenome coverage value for *Cuon alpinus* sample, which is the first result of that quality for the Siberian individual.

Funding: The study is supported by the financial resources of the Russian Scientific Foundation, project No. 23-74-10060.

References

1. Yang D.Y. et al. Technical note: Improved DNA extraction from ancient bones using silica- based spin columns. *Am J Phys Anthropol*. 1998;105(4):539-543. doi 10.1002/(SICI)1096-8644(199804)105:4<539::AID-AJPA10>3.0.CO;2-1
2. Vorobieva N.V., Makunin A.I. et al. High genetic diversity of ancient horses from the Ukok Plateau. *PLoS One*. 2020;15(11):e0241997. doi 10.1371/journal.pone.0241997
3. Maricic T., Whitten M., Pääbo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One.* 2010;5(11):e14004. doi 10.1371/journal.pone.0014004
4. Sanderson C., Radley K., Mayton L. Ethylenediaminetetraacetic acid in ammonium hydroxide for reducing decalcification time. *Biotech Histochem*. 1995;70(1):12-18. doi 10.3109/10520299509108310
5. Jónsson H., Ginolhac A., Schubert M., Johnson P.L.F., Orlando L. MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics.* 2013;29(13):1682-1684. doi 10.1093/bioinformatics/btt193
6. Ronquist F.R., Huelsenbeck J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003;19(12):1572-1574. doi 10.1093/bioinformatics/btg180

5 Симпозиум «Эволюционная, популяционная геномика/генетика и молекулярная филогения: компьютерные и экспериментальные подходы»

Symposium “Evolutionary, population genomics/genetics and molecular phylogeny: computational and experimental approaches”

doi 10.18699/bgrs2024-5.2-01

Исследование разнообразия вирусов в экосистеме озера Байкал с помощью метагеномного анализа

Букин Ю.С.[1, 2]*, Антипова А.Ю.[2]

[1] *Лимнологический институт СО РАН, Иркутск, Россия*

[2] *Иркутский государственный университет, Иркутск, Россия*

* *bukinyira@mail.ru*

Ключевые слова: метагеномика; вирусы; метавиромный анализ; высокопроизводительное секвенирование; озеро Байкал

Введение: Вирусы являются важным компонентом водных экосистем. Концентрация вирусоподобных частиц в пресноводных водоемах достигает 10^{11} на миллилитр. Большинство вирусов являются бактериофагами, поражающими различные виды прокариот. Часть вирусов поражает автотрофный и гетеротрофный планктон, регулируя его численность и динамику. Вирусы также могут поражать любые водные организмы и ассоциированный с ними микробиом. Одним из наиболее перспективных методов исследования вирусов является шотган метагеномный анализ. Данный метод основан на массовом параллельном секвенировании ДНК или РНК из природной смеси организмов с последующей полногеномной сборкой и идентификацией фрагментов геномов или полных геномов вирусов. Оценка покрытия фрагментов вирусных геномов позволяет получить данные об их представленности в исследуемых образцах. К предварительной подготовке генетического материала для исследования вирусов в водных экосистемах существует два подхода. Первый подход предполагает обогащение образца воды путем фильтрации и центрифугирования вирусными частицами. В результате мы получаем массив виромных данных. Второй подход предполагает расшифровку всего генетического материала ДНК или РНК (метагеномный подход) с дальнейшим поиском в общих данных вирусных фрагментов. Такая методика чаше всего применяется для исследования вирусов, ассоциированных с различными водными многоклеточными организмами.

Цель нашего исследования заключалась в анализе таксономического и функционального разнообразия ДНК вирусов в воде озера Байкал и разнообразия ДНК и РНК вирусов, ассоциированных с байкальскими ракообразными амфиподами.

Методы и алгоритмы: Первичные ДНК данные для образцов воды были получены путем обогащения материала вирусными частицами. Для анализа ДНК и РНК вирусов, ассоциированных с байкальскими амфиподами, использовались данные транскриптомного анализа. Данные транкскриптомики позволяют получить представления о разнообразии РНК вирусов и экспрессирующихся фрагментов ДНК вирусов. Расшифровку генетического материала проводили методом дробовика по технологии Illumina MiSeq и Illumina HiSeq.

De novo геномные сборки проводили метагеномным сборщиком metaSPAdes. Идентификация вирусных скаффолдов (фрагментов вирусных геномов) проводилась комбинацией методов, включая методы, основанные на машинном обучении (скрытые Марковские модели) (программы Virsorter, VIBRANT) и методы, основанные на BLAST подобных алгоритмах выравнивания (DIAMOND,

Kaiju). Основой для идентификации вирусов послужили полные вирусные протеомы базы данных NCBI RefSeq. Спектр потенциальных хозяев вирусов определялся по базе данных Virus-Host Database. Полное описание методов можно найти в наших публикациях [1–5].
Результаты: Полное описание результатов по исследованию разнообразия вирусов в образцах воды озера Байкал можно найти в наших публикациях [1, 2]. Эти исследования показали, что в Байкале среди ДНК вирусов доминируют разнообразные бактериофаги, также присутствуют вирусы одноклеточных водорослей. В свободном виде обнаружены вирусные частицы позвоночных и беспозвоночных животных.
В транскриптомах амфипод в большом количестве были представлены экспрессирующиеся гены различных ДНК бактериофагов. Очевидно, что это вирусы, ассоциированные с микробиомом амфипод, и частично вирусы из воды. Также были обнаружены фрагменты РНК вирусов, предсказанными хозяевами для которых были различные беспозвоночные животные, в том числе ракообразные.
Финансирование: Работа поддержана темой бюджетного финансирования ЛИН СО РАН «Генетика сообществ байкальских организмов: структура генофонда, стратегии консервации» 0279-2021-0010 (121032300196-8).

Studying the of viral diversity in the Lake Baikal ecosystem using metagenomic analysis

Bukin Yu.S.[1, 2]*, Antipova A.Yu.[2]
[1] *Limnological Institute, SB RAS, Irkutsk, Russia*
[2] *Irkutsk State University, Irkutsk, Russia*
* *bukinyira@mail.ru*

Key words: metagenomics; viruses; metavirome analysis; high-throughput sequencing; Lake Baikal

Motivation and Aim: Viruses are an important component of aquatic ecosystems. The concentration of virus-like particles in freshwater bodies reaches up to 10^{11} per milliliter. Most viruses are bacteriophages that infect various types of prokaryotes. Some viruses infect autotrophic and heterotrophic plankton, regulating their numbers and dynamics. Viruses can also infect any aquatic organism and its associated microbiome.
One of the most promising methods for studying viruses is shotgun metagenomic analysis. This method is based on massively parallel sequencing of DNA or RNA from a natural mixture of organisms, followed by whole-genome assembly and identification of genome fragments or complete genomes of viruses. Assessing the coverage of viral genome fragments allows us to obtain data on their representation in the studied samples. There are two approaches to the preliminary preparation of genetic material for the study of viruses in aquatic ecosystems. The first approach involves enriching a water sample by viral particles with filtration and centrifugation. As a result, we obtain an array of virome data. The second approach involves decoding the entire genetic material of DNA or RNA (metagenomic approach) with a further search for viral fragments in the general data. This technique is most often used to study viruses associated with various aquatic multicellular organisms.

The purpose of our study was to analyze the taxonomic and functional diversity of DNA viruses in the water of Lake Baikal and the diversity of DNA and RNA viruses associated with Baikal amphipod.

Methods and Algorithms: Primary DNA data for water samples was obtained by enriching the material with viral particles. Transcriptomic analysis data were used to analyze DNA and RNA viruses associated with Baikal amphipods. Transcriptomics data provide insight into the diversity of RNA viruses and expressed DNA viruses fragments. Genetic material was sequenced using the shotgun method using Illumina MiSeq and Illumina HiSeq technology.

De novo genome assemblies were performed with the metaSPAdes metagenomic assembler. Identification of viral scaffolds (fragments of viral genomes) was carried out by a combination of methods, including methods based on machine learning (hidden Markov models) ("Virsorter" and "VIBRANT" programs) and methods based on BLAST-like alignment algorithms ("DIAMOND", "Kaiju") The basis for virus identification was the complete viral proteomes of the NCBI RefSeq database. The range of potential virus hosts was determined using the Virus-Host Database. A complete description of the methods can be found in our publications [1–5].

Results: A complete description of the results of studying the diversity of viruses in water samples from Lake Baikal can be found in our publications [1, 2]. These studies showed that in Baikal, various bacteriophages dominate among DNA viruses; viruses of unicellular algae are also present. Viral particles of vertebrate and invertebrate animals were found in free form.

Expressed genes of various DNA bacteriophages were abundantly represented in amphipod transcriptomes. Obviously, these are viruses associated with the microbiome of amphipods and partly viruses from water. Fragments of RNA viruses were also discovered, the predicted hosts for which were various invertebrate animals, including crustaceans.

Funding: Work supported by budget funding of Limnological Institute SB RAS "Genetics of communities of Baikal organisms: structure of the gene pool, conservation strategies" 0279-2021-0010 (121032300196-8).

Список литературы/References

1. Butina T.V., Bukin Y.S., Krasnopeev A.S., Belykh O.I., Tupikin A.E., Kabilov M.R., Belikov S.I. Estimate of the diversity of viral and bacterial assemblage in the coastal water of Lake Baikal. *FEMS Microbiol Lett.* 2019;366(9):fnz094. doi 10.1093/femsle/fnz094
2. Butina T.V., Bukin Y.S., Petrushin I.S., Tupikin A.E., Kabilov M.R., Belikov S.I. Extended evaluation of viral diversity in lake baikal through metagenomics. *Microorganisms*. 2021;9(4):760. doi 10.3390/microorganisms9040760
3. Butina T.V., Petrushin I.S., Khanaev I.V., Bukin Y.S. Metagenomic assessment of DNA viral diversity in freshwater sponges, Baikalospongia bacillifera. *Microorganisms.* 2022;10(2):480. doi 10.3390/microorganisms10020480
4. Alexyuk M.S., Bukin Y.S., Butina T.V., Alexyuk P.G., Berezin V.E., Bogoyavlenskiy A.P. Viromes of coastal waters of the North Caspian Sea: Initial assessment of diversity and functional potential. *Diversity*. 2023;15(7):813. doi 10.3390/d15070813
5. Bukin Y.S., Bondaryuk A.N., Butina T.V. Performance analysis of cross-assembly of metatranscriptomic datasets in viral community studies. *Math Biol Bioinf.* 2023;18(2):418-433. doi 10.17537/2023.18.418

doi 10.18699/bgrs2024-5.2-02

Разнообразие вирусов в образцах байкальских брюхоногих моллюсков (*Benedictia baicalensis*)

Бутина Т.[1]*, Букин Ю.[1], Бондарюк А.[2], Петрушин И.[3], Ханаев И.[1]

[1] *Лимнологический институт СО РАН, Иркутск, Россия*

[2] *Иркутский научно-исследовательский противочумный институт Сибири и Дальнего Востока, Иркутск, Россия*

[3] *Иркутский государственный университет, Иркутск, Россия*

* *tvbutina@mail.ru*

Ключевые слова: вирусы; вирусное разнообразие; метатранскриптомный анализ; пресноводные брюхоногие моллюски; озеро Байкал

Мотивация и цель: Первые сведения о наличии вирусов у моллюсков появились в начале 1970-х гг. на основе гистологических и цитологических наблюдений с использованием световой и электронной микроскопии [1]. В настоящее время в связи с проблемой появления новых вирусных заболеваний и расширением возможностей молекулярной вирусологии реализуются проекты по массовому параллельному секвенированию (в основном метатранскриптомные исследования) различных видов беспозвоночных, в том числе моллюсков [2, 3]. Проведенные исследования выявили огромное разнообразие по большей части неизвестных вирусов и заполнили пробелы в истории эволюции вирусов. Большинство исследований сосредоточено на морских, преимущественно двустворчатых моллюсках, и значительно реже на брюхоногих моллюсках. Гораздо меньше внимания уделяется пресноводным моллюскам. В нашей работе мы провели оценку вирусного разнообразия в образцах байкальских эндемичных брюхоногих моллюсков *Benedictia baicalensis* с помощью метатранскриптомного анализа.

Методы и алгоритмы: Моллюски рода *Benedictia* (*B. baicalensis*) были отобраны на двух станциях Южного Байкала: в районе пос. Листвянка и пос. Большие Коты; а также в Северном Байкале вблизи Ушканьих островов. Отбор проб осуществлялся водолазами, как описано ранее [4].

Суммарную РНК выделяли из образцов *B. baicalensis* с использованием TRI Reagent (Molecular Research Center, США) и набора Direct-zol RNA MiniPrep (Zymo Research, США). Библиотеки готовили с использованием набора MGI Easy RNA Library Prep Kit (MGI Tech Co., Ltd., Китай) в соответствии с протоколами производителя. Библиотеки кДНК секвенировали с обоих концов на платформе DNBSEQTM с использованием секвенатора DNBSEQ-400 и набора для высокопроизводительного секвенирования DNBSEQ-G400RS (MGI Tech Co., Ltd., Китай). Длина секвенированных фрагментов составляла 150 оснований.

Проверку качества наборов данных (парных прочтений) осуществляли с помощью программы FASTQC v.0.11.9. Фильтрацию прочтений по качеству проводили с помощью программы Trimmomatic v.0.32 (MAXINFO:40:0.05 AVGQUAL:15 MINLEN:90). После обработки наборы данных содержали от 63 до 67 млн пар прочтений на образец.

Метатранскриптомные данные были агрегированы в один массив для комбинированной («микс») сборки *de novo* [5] с помощью программы MEGAHIT v.1.2.9. С использованием программы VirSorter2 v.2.2.4 были идентифицированы вирусные скаффолды и открытые рамки считывания (ORFs). Результаты идентификации вирусов проверялись программой CheckV. Все скаффолды, идентифицированные VirSorter2 как вирусные, сравнивались с хромосомными геномами моллюсков и протеомными данными из базы данных NCBI с использованием алгоритма BLASTn и инструмента DIAMOND v2.0.14.152. Вирусные скаффолды также проверялись на наличие мобильных генетических элементов с помощью программы DFAMSCAN.

Таксономическую идентификацию вирусных скаффолдов осуществляли путем сравнения предсказанных вирусных белков с базой данных полных вирусных протеомов NCBI RefSeq с помощью алгоритма DIAMOND. Программы Bowtie2 v.1.3.1 и SAMtools v.1.7 использовались для определения количества ридов из каждого образца, входящих в состав вирусных скаффолдов. Предсказанные вирусные ORFs сопоставлялись с функциональными мотивами из базы данных Pfam с использованием программного обеспечения PfamScan v.1.5 и базы данных консервативных доменов CDD (Conserved Domain Database).

Филогенетическую реконструкцию на основе генов *RdRp* проводили с помощью IQTREE v.1.6.12. В наш анализ были включены BLAST-хиты из базы NCBI nr и известные вирусы из базы данных RefSeq с наибольшей степенью сходства аминокислотных последовательностей с идентифицированными нами вирусами. Выравнивание аминокислотных последовательностей проводили с помощью MAFFT v.7.490. Наиболее подходящая модель замены аминокислот (LG+G4+Inv+F) была выбрана на основе значений BIC критерия, с помощью ModelFinder, реализованного в программе IQTREE.

Результаты: В метатранскриптомных наборах данных из образцов байкальских моллюсков было выявлено 64 вирусных скаффолда; из них только один был идентифицирован как ДНК-вирус, родственный одноцепочечным ДНК CRESS-вирусам из пресноводных водоемов Новой Зеландии (озера Сара) и США. Остальные 63 скаффолда (от 24 до 33 на образец) были идентифицированы как РНК-вирусы, сходные с таковыми из разных семейств вирусов (*Picornaviridae*, *Marnaviridae*, *Dicistroviridae*, *Solemoviridae*, *Partitiviridae*, *Tombusviridae*, *Nodaviridae*, *Narnaviridae*, *Caulimoviridae* и *Qinviridae*). В целом четыре исследованные пробы были схожими, несмотря на большие расстояния между станциями отбора проб. Это можно объяснить активными процессами перемешивания вод озера Байкал, в том числе за счет течений и ветров [6], а также длительной циркуляцией разнообразных вирусов в экосистеме озера.

Среди идентифицированных вирусных геномов (скаффолдов) оказались пикорноподобные вирусы, родственные ранее описанным вирусам пресноводных моллюсков, *Biomphalaria virus 1* и *Biomphalaria virus 3*. Как известно, пикорноподобные вирусы могут вызывать массовую смертность морских моллюсков [7].

В целом сходство белков вирусов из *B. baicalensis* с известными вирусами из базы данных NCBI было низким и в единичных случаях превышало 70 %. Только один скаффолд имел 99.3 % сходства (при выравнивании 99 % последовательности) с неструктурированным полипротеином вируса некроза тигрового фугу (*Tiger puffer nervous necrosis virus*, семейство *Nodaviridae*). Как оказалось, резервуарами «вирусов нервного некроза» могут быть различные двустворчатые, брюхоногие и

головоногие моллюски, а также другие виды морских беспозвоночных (крабы, креветки, артемии и коловратки), обнаруженные в водах Южной Кореи, Японии и Китая, в европейской части Атлантического океана и Средиземном море [8]. Скорее всего, байкальские моллюски также могут выступать резервуарами для нодавирусов, как и других вирусов, и влиять на их эпидемиологию.
На филогенетическом древе последовательности *RdRp* из образцов байкальских брюхоногих моллюсков вошли в состав пяти кластеров, соответствующих РНК-вирусам порядков *Picornavirales*, *Tolivirales*, *Nodamuvirales*, *Durnavirales* и *Sobelivirales.* Полученные результаты демонстрируют отдаленное родство выявленных нами вирусных геномов (геномных фрагментов), а также разнообразие РНК-вирусов с широким спектром хозяев (моллюски, насекомые, растения, водоросли, простейшие, грибы и позвоночные животные). Наши данные выявили также значительные генетические расстояния между генами *RdRp* байкальских вирусов и вирусов из других организмов и биотопов, что указывает на уникальность обнаруженных нами вирусов.
Выводы: Проведенное метатранскриптомное исследование образцов байкальских моллюсков позволило раскрыть новый спектр уникальных вирусов в экосистеме озера Байкал. Выявленное высокое разнообразие вирусов в байкальских эндемичных брюхоногих моллюсках *B. baicalensis* включало неизвестных представителей разных отрядов и семейств, поражающих широкий круг хозяев.
B. baicalensis обитают на каменистых и смешанных песчано-каменистых биотопах и являются всеядным видом, питающимся субстратом [9]. Это способствует передаче и биоаккумуляции различных вирусов, поражающих широкий круг придонных и планктонных обитателей озера Байкал. Таким образом, для установления фактического круга хозяев новых вирусов и выявления истинного вирома байкальских моллюсков необходимо продолжать начатые нами исследования.
Финансирование: Исследование поддержано Российским научным фондом (№ 22-24-01120) и Министерством науки и высшего образования РФ (№ 0279-2021-0005).

Viral diversity in samples of Baikal gastropods (*Benedictia baicalensis*)

Butina T.[1]*, Bukin Y.[1], Bondaryuk A.[2], Petrushin I.[3], Khanaev I.[1]

[1] *Limnological Institute, SB RAS, Irkutsk, Russia*

[2] *Irkutsk Anti-Plague Research Institute of Siberia and Far East, Irkutsk, Russia*

[3] *Irkutsk State University, Irkutsk, Russia*

** tvbutina@mail.ru*

Key words: viruses; viral diversity; metatranscriptomic analysis; freshwater gastropods; Lake Baikal

Motivation and Aim: The first evidence of viruses in mollusks appeared in the early 1970s on the basis of histological and cytological observations using light and electron microscopy [1]. Currently, due to the problem of the emergence of new viral diseases and the expansion of the capabilities of molecular virology, the different projects are being implemented for massive parallel sequencing (mainly metatranscriptomic studies) of various species of invertebrate, including mollusks [2, 3]. These studies have revealed

a huge diversity of mostly previously unknown viruses and filled gaps in the evolutionary history of viruses. Most studies have primarily focused on marine mollusks, mostly bivalves and less often gastropods. Much less attention has been paid to freshwater mollusks. In our study, we estimated the viral diversity in samples of the Baikal endemic gastropod *Benedictia baicalensis* using metatranscriptomic analysis (total RNA-sequencing).

Methods and Algorithms: Mollusks of the genus *Benedictia* (*Benedictia baicalensis*) were taken at two stations of Southern Baikal: in Listvennichny Bay (near the village of Listvyanka) and near the village of Bolshie Koty; and near the Ushkany Islands in Northern Baikal. Sampling was carried out by divers [4].

Total RNA was isolated from samples *B. baicalensis* (pools of five) using TRI Reagent (Molecular Research Center, USA) and a Direct-zol RNA MiniPrep kit (Zymo Research, USA). Libraries were prepared using the MGI Easy RNA Library Prep Kit (MGI Tech Co., Ltd., China) in accordance with the manufacturer's protocols. The cDNA libraries were sequenced from both ends on a DNBSEQTM platform using a DNBSEQ-400 NGS sequencer and a DNBSEQ-G400RS High-throughput Sequencing Kit (MGI Tech Co., Ltd., Wuhan, China). The length of the sequenced fragments was 150 bases.

The quality visualization of datasets (paired reads) was carried out using the FASTQC v.0.11.9 program. Trimming of reads by quality was carried out with the Trimmomatic v.0.32 program (MAXINFO:40:0.05 AVGQUAL:15 MINLEN:90). After trimming, the datasets finally contained 63–67 million read pairs.

All metatranscriptomic data were aggregated into one array for *de novo* cross-assembly [5] using the MEGAHIT v.1.2.9 NGS assembler. We identified the viral scaffolds and open reading frames (ORFs) within them using the VirSorter2 v.2.2.4 tool. The results of virus identification were checked by the CheckV program. All scaffolds identified as viral by VirSorter2 were compared with chromosomal genome and proteomic data of mollusks from the NCBI database using BLASTn algorithm and DIAMOND v2.0.14.152 tool. The viral scaffolds were also checked for mobile genetic elements using the DFAMSCAN program.

Taxonomic identification for the viral scaffolds was carried out by comparisons of predicted viral proteins in scaffolds with the NCBI RefSeq complete viral proteome database and DIAMOND algorithm. The Bowtie2 v.1.3.1 (with default options) and SAMtools v. 1.7 results were used to determine the number of reads mapped on each predicted viral scaffold from each sample. The predicted viral proteins (ORFs) were matched with functional motifs in the Pfam database using PfamScan v.1.5 software and the Conserved Domain Database.

Phylogenetic reconstruction based on *RdRp* genes was performed with IQTREE v.1.6.12. Our analysis included NCBI nr BLAST-hits and known, well-studied viruses from the RefSeq database with a significant degree of amino acid sequence similarity to the viruses we identified. Aligning of amino acid sequences was carried out in MAFFT v.7.490. The best-fitting amino acid substitution model (LG+G4+Inv+F) was selected based on the BIC criterion values calculated by ModelFinder implemented in IQTREE.

Results: In total in metatranscriptomic data from the samples of Baikal mollusks we revealed 64 viral scaffolds; of these, only one was identified as a DNA virus related to ssDNA CRESS viruses from a New Zealand freshwater Lake Sarah and USA freshwater springs. The rest 63 viral scaffolds (from 24 to 33 per sample) were identified as RNA viruses similar to those from different viral families (*Picornaviridae*, *Marnaviridae*, *Dicistroviridae*, *Solemoviridae*, *Partitiviridae*, *Tombusviridae*, *Nodaviridae*,

Narnaviridae, *Caulimoviridae* and *Qinviridae*). Overall, the four examined pool samples were similar despite the large distances between sampling stations. This can be explained by the active mixing processes of the Lake Baikal waters, including those due to the currents and winds [6], as well as the long-term circulation of a wide variety of viruses in the lake's ecosystem.

Among the viral genomes (or scaffolds) identified were picorna-like viruses closely related to the previously described freshwater shellfish viruses *Biomphalaria virus 1* and *Biomphalaria virus 3*. As known, picorna-like viruses may also be responsible for mass mortality events in marine mollusks [7].

In general, the similarity of proteins of *B. baicalensis* viruses to known viruses from the NCBI database was low and in single cases exceeded 70 %. Only one scaffold showed 99.3 % identity and 99 % query coverage with unstructured polyprotein of *Tiger puffer nervous necrosis virus* (TPNNV, family *Nodaviridae*). Reservoirs of 'nervous necrosis viruses' were reported in different mollusks (bivalves, gastropods and cephalopods) and other marine invertebrate species (crabs, shrimps, artemias and rotifers) in waters of South Korea, Japan and China, in the European Atlantic and Mediterranean waters [8]. Baikal mollusks may also act as reservoirs for nodaviruses and other viruses and influence their epidemiology.

RdRp sequences from Baikal gastropods were distributed into five clusters corresponding to the RNA viral orders *Picornavirales*, *Tolivirales*, *Nodamuvirales*, *Durnavirales* and *Sobelivirales*, demonstrating the distant relationship of the identified viral genomes (genomic fragments) and the diversity of RNA viruses with a wide range of hosts (mollusks, insects, plants, algae, protozoa, fungi and vertebrates). Our data also revealed significant distances between the *RdRps* of Baikal viruses and those from other organisms and biotopes, indicating the novelty of the viruses we discovered.

Conclusion: The metatranscriptomic study of samples of Baikal mollusks made it possible to reveal a new range of unique viruses in the ecosystem of Lake Baikal. The detected high diversity of viruses in the Baikal endemic gastropod *B. baicalensis* included unknown representatives of different orders and families infecting a wide range of hosts.

B. baicalensis lives on rocky and mixed sandy-stony biotopes and is an omnivorous species that grazes on the substrate [9]. This contributes to the transmission and bioaccumulation of various viruses that infect a wide range of benthic and planktonic inhabitants of Lake Baikal. Thus, additional research is required to establish the actual host range for novel viruses and uncover the true virome of Baikal mollusks.

Funding: The study is supported by the Russian Science Foundation (No. 22-24-01120) and the Ministry of Science and Higher Education of the Russian Federation (No. 0279-2021-0005).

Список литературы/References

1. Farley C.A. Viruses and virus-like lesions in marine mollusks. *Mar Fish Rev.* 1978;40:18-20
2. Shi M. et al. Redefining the invertebrate RNA virosphere. *Nature*. 2016;540:539-543. doi 10.1038/nature20167
3. Zhang Y.Y., Chen Y., Wei X., Cui J. Viromes in marine ecosystems reveal remarkable invertebrate RNA virus diversity. *Sci China Life Sci.* 2022;65:426-437. doi 10.1007/s11427-020-1936-2
4. Butina T.V. et al. Viral diversity in samples of freshwater gastropods *Benedictia baicalensis* (caenogastropoda: benedictiidae) revealed by total RNA-sequencing. *Int J Mol Sci.* 2023;24:17022. doi 10.17537/2023.18.418
5. Букин Ю.С., Бондарюк А.Н., Бутина Т.В. Анализ эффективности микс-сборки метатранскриптомных наборов данных в исследовании вирусных сообществ. *Математическая биология и биоинформатика.* 2023;18(2):418-433. doi 10.17537/2023.18.418

[Bukin Y.S., Bondaryuk A.N., Butina T.V. Performance analysis of cross-assembly of metatranscriptomic datasets in viral community studies. *Math Biol Bioinf.* 2023;18:418-433 (in Russian)]

6. Байкал: атлас. Москва: Роскартография, 1993
[Baikal: atlas. Moscow: Roscartography, 1993 (in Russian)]
Renault T., Novoa B. Viruses infecting bivalve molluscs. *Aquat Living Resour.* 2004;17:397-409. doi 10.1051/alr:2004049
7. Johnstone C., Pérez M., Arizcun M., García-Ruiz C., Chaves-Pozo E. Reservoirs of red-spotted grouper nervous necrosis virus (RGNNV) in squid and shrimp species of Northern Alboran Sea. *Viruses*. 2022;14:328. doi 10.3390/v14020328
8. Ситникова Т.Я., Репсторф П. Эти моллюски живут только в Байкале. *Наука из первых рук*. 2004;1(2):84-99
[Sitnikova T.Y., Roepstorf P. Mollusks that Live Exclusively in Lake Baikal. *Nauka iz Pervykh Ruk*. 2004;1(2):84-99 (in Russian)]

doi 10.18699/bgrs2024-5.2-03

Вклад методов молекулярной филогенетики в контроль разнообразия микробиома легких у пациентов с муковисцидозом при проведении комплексной терапии

Воронина О.Л.[1]*, Рыжова Н.Н.[1], Кунда М.С.[1], Ермолова Е.И.[1], Кагазежев Р.У.[2], Амелина Е.Л.[2], Гинцбург А.Л.[1]

[1] *ФГБУ НИЦЭМ им. Н.Ф. Гамалеи Минздрава России, Москва, Россия*

[2] *ФГБУ НИИ пульмонологии ФМБА России, Москва, Россия*

* *olv550@gmail.com*

Ключевые слова: микробиом; *Pseudomonadota*; муковисцидоз; филогенетическое разнообразие; биомаркеры микробиома

Мотивация и цель: Применение методов секвенирования ДНК в изучении разнообразия микробиоты различных биотопов человеческого организма, создание баз данных и программ для аннотирования полученных прочтений позволили охарактеризовать разнообразные микробиоценозы условно здоровых людей [1]. При заболевании происходят изменения в составе микробного сообщества. Для пациентов с муковисцидозом (МВ) наиболее важным отслеживаемым биотопом является респираторный тракт, главным образом нижние дыхательные пути. С 2016 г. в терапии российских пациентов с МВ начали использовать модуляторы CFTR (cystic fibrosis transmembrane conductance regulator), способствующие фолдингу и поддержанию в активном состоянии белка хлорного канала, что нормализует функции эпителиоцитов и приводит к улучшению общего состояния пациентов [2]. Для хронически инфицированных пациентов путь восстановления микробиоты респираторного тракта долог и требует подбора комплексной терапии. В анализе эффективности протокола лечения помогает сравнение микробиомов мокроты на этапах терапии. Ключевыми в сравнительном исследовании образцов являются методы молекулярной филогенетики и основанные на них филогенетические протоколы [3]. Мы применили такой протокол для многофакторного исследования образцов взрослых пациентов с МВ, хронически инфицированных клиническими значимыми *Pseudomonadota*, для определения эффективности комплексной терапии, включающей антибактериальные и таргетные препараты.

Методы и алгоритмы: Выборка включала 103 образца мокроты от 52 взрослых пациентов с МВ, из которых 56 % составили женщины и 44 % – мужчины. Образцы, собранные в течение 11 месяцев, группировали по следующим параметрам: возраст пациентов (18–23, 24–30, 31–56); мутации в гене *cftr* (F508del/F508del, F508del/неF508de, неF508del/неF508del); фенотип (мягкий, тяжелый); степень тяжести болезни легких: легкое течение (mild lung disease, MiLD, ОФВ1 ≥ 70 %), средней тяжести (moderate lung disease, MoLD, 40 % ≤ ОФВ1 < 70 %), тяжелое течение (severe lung disease, SLD, ОФВ1 < 40 %); клиническое состояние (вне обострения, baseline, B; обострение, exacerbation, E; эпизодическая антибиотикотерапия, treatment, AT), инфицирующие микроорганизмы (PCH – хронически инфицированные клинически значимыми *Pseudomonadota*; P – с появляющимися

клинически значимыми *Pseudomonadota*; NP – без клинически значимых *Pseudomonadota* в микробиоме легких); таргетная терапия (наличие – TT, отсутствие – NTT).

Тотальную ДНК из образцов выделяли с помощью MagAttract HMW DNA Kit (QIAGEN, США). ДНК анализировали с помощью экспресс-диагностики, включающей амплификацию и секвенирование по Сэнгеру следующих мишеней: *16S rDNA*, *trpE – P. aeruginosa*, *gltB*, *gyrB – Burkholderia*, *nrdA – Achromobacter*, *adkA – E. coli*, *rpoB – K. pneumonia*, а также с помощью массового параллельного секвенирования ампликонов *16S rDNA* (V1-V5). Результаты депонировали в GenBank (биопроект PRJNA717158). Для анализа прочтений использовали Microbial Genomics Module пакета CLC Genomic Workbench v.21.0.1. (QIAGEN, США), для определения операционных таксономических единиц (Operational Taxonomic Unit, OTU) – базу данных Greengenes v13_8 с уровнем сходства 97 %. Для оценки альфа-разнообразия применяли индексы Shannon entropy и коэффициент филогенетического разнообразия (PD). Достоверность отличия альфа-разнообразия по группам образцов рассчитывали с помощью H-критерия Крускала–Уоллиса и U-критерия Манна–Уитни, а также подтверждали анализом PERMANOVA (p-Value < 0.05). Бета-разнообразие оценивали с помощью индекса Euclidean distance, а также метрик UniFrac. Для подробного анализа сходства между образцами использовали Principal Coordinate Analysis (PCoA) и статистические функции MS Exel. Для обнаружения таксонов, являющихся маркерами различий между анализируемыми группами, использовали алгоритм LEfSe (Linear discriminant analysis Effect Size) [4]. В расчет брали только достоверные значения LDA Score (log 10) >4–5.

Результаты: Анализ всей выборки образцов показал отсутствие зависимости филогенетического разнообразия (PD) от фенотипа мутации и наличия/отсутствия таргетной терапии. Вместе с тем PD достоверно снижалось в образцах РСН пациентов по сравнению с P и NP группами, что было характерно как для наиболее массовой подгруппы инфицированных *P. aeruginosa*, так и для подгрупп с *Burkholderia s*pp., *Achromobacter* spp. и *E. coli*. Наиболее низкие показатели PD наблюдали в образцах SLD пациентов, достоверно отличавшиеся от образцов MoLD и MiLD групп. Сравнение PD образцов в разных состояниях пациентов выявило наиболее высокие показатели вне обострения и достоверно более низкие в состояниях E и AT. Оценка бета-разнообразия с помощью индекса Euclidean distance продемонстрировала группировку образцов по основным возбудителям инфекции с миграцией в направлении групп P и NP образцов РСН *P. aeruginosa* при доле *Staphylococcus* 9–48 %. Использование метрики «взвешенный не нормализованный UniFrac», учитывающей значимость редких таксонов, позволило выявить больше мигрирующих РСН образцов, в том числе тех, в которых нарастала доля *Streptococcus* (45 %) или *Capnocytophaga* (41 %) на этапах терапии. Алгоритм LEfSe применяли для выявления маркерных микроорганизмов для групп многофакторного анализа. Так, образцы SLD пациентов отличали *Burkholderiales*, а группы MoLD и MiLD – представители *Actinomycetota* и *Bacillota* соответственно. Маркерами образцов, сгруппированных по мутациям пациентов, были *Bacillota* при мутации F508del/F508del, *Pseudomonas* при F508del/неF508del и *Enterobacteriaceae* при неF508del/неF508del. Маркерами образцов в состоянии вне обострения были *Streptococcus* и *Bacteroidota* (прежде всего, *Prevotella*), в обострении – *Flavobacteriia* и *Capnocytophaga*, после

эпизодической антибиотикотерапии – *Bacilli* (*Staphylococcus*) и *Achromobacter*. В группах образцов по времени приема таргетного препарата *Burkholderia* была маркером двух лет терапии, а *Achromobacter* – 5-летнего лечения модуляторами CFTR. Таким образом, сокращение доли *Burkholderiales* происходит очень медленно даже при длительной таргетной терапии на фоне поддерживающей и эпизодической антибиотикотерапии. Таргетная терапия не исключает обострений, после которых доля *Burkholderiales* возрастает. Снижение доли другого представителя *Pseudomonadota*, *P. aeruginosa*, приводит к другой проблеме – росту доли *Staphylococcus*, конкурента *P. aeruginosa* [5]. Анализ образцов показал, что PD сохраняется на высоком уровне при 4–29 % *Staphylococcus* в микробиоме и достоверно снижается в образцах как с низкой долей *Staphylococcus* (<4 %), так и с высокой (>30 %). При восстановлении микробиома наблюдается дисбаланс для многих комменсалов ротовой полости, поступающих в нижние дыхательные пути посредством микроаспирации, например, *Streptococcus*, *Prevotella*, *Rothia.* Предполагается роль этих бактерий и их метаболитов в воспалительных процессах в легких [6].

Выводы: Исследования показали, что возвращение сбалансированного разнообразия микробиома легких у хронически инфицированных взрослых пациентов с МВ возможно при длительной комплексной терапии и внимании к представителям всех таксонов, составляющих биотоп нижних дыхательных путей, выявить которые позволяют методы молекулярной филогенетики.

Финансирование: Исследование поддержано Государственным заданием НИЦЭМ им. Н.Ф. Гамалеи (№ 056-00066-23-00).

The contribution of molecular phylogenetic methods to the control of the diversity of the lung microbiome in patients with cystic fibrosis during complex therapy

Voronina O.[1]*, Ryzhova N.[1], Kunda M.[1], Ermolova E.[1], Kagazezhev R.[2], Amelina E.[2], Gintsburg A.[1]

[1] *N.F. Gamaleya National Research Center for Epidemiology and Microbiology, Ministry of Health of Russia, Moscow, Russia*

[2] *Pulmonology Research Institute under FMBA of Russia, Moscow, Russia*

* *olv550@gmail.com*

Key words: microbiome; *Pseudomonadota*; cystic fibrosis; phylogenetic diversity; microbiome biomarkers

Motivation and Aim: The use of DNA sequencing methods in studying the diversity of the microbiota of various biotopes of the human body, the creation of databases and programs for annotating the obtained reads made it possible to characterize the various microbiocenoses of relatively healthy people [1]. During disease, changes occur in the composition of the microbial community. For patients with cystic fibrosis (CF), the most important biotope monitored is the respiratory tract, mainly the lower respiratory tract. Since 2016, CFTR (cystic fibrosis transmembrane conductance regulator) modulators have been used in the treatment of Russian patients with CF, which promote the folding and maintenance of the chloride channel protein in an active state, which normalizes the functions of epithelial cells and leads to an improvement in the general condition of

patients [2]. For chronically infected patients, the path to restoring the microbiota of the respiratory tract is long and requires the selection of complex therapy. Comparison of sputum microbiomes at therapy stages helps in analyzing the effectiveness of a treatment protocol. The key to the comparative study of samples are the methods of molecular phylogenetics and phylogenetic protocols based on them [3]. We applied such a protocol to a multifactorial study of samples from adult CF patients chronically infected with clinically significant *Pseudomonadota* to determine the effectiveness of combination therapy, including antibacterial and targeted drugs.

Methods and Algorithms: 103 sputum samples from 52 adult patients with CF, of whom 56 % were women and 44 % were men, were included in the analysis. Samples collected over 11 months were grouped according to the following parameters: patient age (18–23, 24–30, 31–56); mutations in the *cftr* gene (F508del/F508del, F508del/non-F508del, non-F508del/non-F508del); phenotype (mild, severe); severity of lung disease: mild (mild lung disease, MiLD, FEV1 ≥ 70 %), moderate (moderate lung disease, MoLD, 40 % ≤ FEV1 < 70 %), severe (severe lung disease, SLD, FEV1 < 40 %); clinical condition (out of exacerbation, baseline, B; exacerbation, E; episodic antibiotic therapy, treatment, AT), infecting microorganisms (PCH – chronically infected with clinically significant Pseudomonadota; P – with emerging clinically significant *Pseudomonadota*; NP – without clinically significant *Pseudomonadota* in lung microbiome); targeted therapy (presence – TT, absence – NTT). Total DNA from the samples was isolated using MagAttract HMW DNA Kit (QIAGEN, USA). DNA was analyzed using express diagnostics, including amplification and Sanger sequencing of the following targets: *16S rDNA*, *trpE – P. aeruginosa*, *gltB*, *gyrB – Burkholderia*, *nrdA – Achromobacter*, *adkA – E. coli*, *rpoB – K. pneumonia*, and using massively parallel sequencing of 16S rDNA amplicons (V1-V5).

The results were deposited in GenBank (bioproject PRJNA717158). Microbial Genomics Module of the CLC Genomic Workbench v.21.0.1 package (QIAGEN, USA) was used for analysis of reads. Greengenes v13_8 database with a similarity level of 97 % was used for determination Operational Taxonomic Units (OTU). Shannon entropy index and the coefficient of phylogenetic diversity (PD) were used to assess alpha diversity. The significance of the difference in alpha diversity between groups of samples was calculated using the Kruskal–Wallis H test and the Mann–Whitney U test, and was also confirmed by PERMANOVA analysis (p-Value < 0.05). Beta diversity was assessed using the Euclidean distance index, as well as UniFrac metrics. For a detailed analysis of the similarity between samples, Principal Coordinate Analysis (PCoA) and MS Exel statistical functions were used. To detect taxa that are markers of differences between the analyzed groups, the LEfSe (Linear discriminant analysis Effect Size) algorithm was used [4]. Only reliable values of LDA Score (log 10) >4–5 were taken into account.

Results: Analysis of the entire sample collection showed no dependence of phylogenetic diversity (PD) on the phenotype of the mutation and the presence/absence of targeted therapy. At the same time, PD significantly decreased in the samples of PCH patients compared to the P and NP groups, which was typical both for the most widespread subgroup of those infected with *P. aeruginosa*, and for the subgroups with *Burkholderia* spp., *Achromobacter* spp. and *E. coli*. The lowest PD rates were observed in the samples of SLD patients, which were significantly different from the samples of the MOLD and MiLD groups. Comparison of PD samples in different patient states revealed the highest rates outside of exacerbation and significantly lower ones in states E and AT.

Assessment of beta diversity using the Euclidean distance index demonstrated grouping of samples according to the main pathogens with migration towards the P and NP groups of PCH *P. aeruginosa* samples with a proportion of *Staphylococcus* of 9–48 %. The use of the "weighted non-normalized UniFrac" metric, which takes into account the significance of rare taxa, made it possible to identify more migrating PCH samples, including those in which the proportion of *Streptococcus* (45 %) or *Capnocytophaga* (41 %) increased during the treatment stages. The LEfSe algorithm was used to identify marker microorganisms for multivariate analysis groups. Thus, the SLD patient samples were distinguished by *Burkholderiales*, and the MoLD and MiLD groups were distinguished by representatives of *Actinomycetota* and *Bacillota*, respectively. Markers for samples grouped by patient mutations were *Bacillota* for F508del/F508del mutation, *Pseudomonas* for F508del/non-F508del, and *Enterobacteriaceae* for non-F508del/non-F508del. The markers of the samples in the non-exacerbation state were *Streptococcus* and *Bacteroidota* (primarily *Prevotella*), in the exacerbation – *Flavobacteriia* and *Capnocytophaga*, after episodic antibiotic therapy – *Bacilli* (*Staphylococcus*) and *Achromobacter*. In groups of samples based on the time of taking the targeted drug, *Burkholderia* was a marker of two years of therapy, and *Achromobacter* was a marker of 5 years of treatment with CFTR modulators. Thus, the reduction in the proportion of *Burkholderiales* occurs very slowly even with long-term targeted therapy against the background of maintenance and episodic antibiotic therapy. Targeted therapy does not exclude exacerbations, after which the proportion of *Burkholderiales* increases. A decrease in the share of another representative of *Pseudomonadota*, *P. aeruginosa*, leads to another problem – an increase in the share of *Staphylococcus*, a competitor of *P. aeruginosa* [5]. Analysis of the samples showed that PD remains at a high level with 4–29 % *Staphylococcus* in the microbiome and is significantly reduced in both samples with a low proportion of *Staphylococcus* (<4 %) and high (>30 %). When the microbiome is restored, an imbalance is observed for many oral commensals entering the lower respiratory tract through microaspiration, for example, *Streptococcus*, *Prevotella*, *Rothia*. The role of these bacteria and their metabolites in inflammatory processes in the lungs is assumed [6].

Conclusion: Studies have shown that the return of balanced diversity of the lung microbiome in chronically infected adult patients with CF is possible with long-term complex therapy and attention to representatives of all taxa that make up the biotope of the lower respiratory tract, which can be identified by molecular phylogenetics methods.

Funding: The study is supported by the State assignment of the N.F. Gamaleya National Research Center for Epidemiology and Microbiology (No. 056-00066-23-00).

Список литературы/References

1. Dickson R.P., Erb-Downward J.R., Freeman C.M., McCloskey L., Beck J.M., Huffnagle G.B., Curtis J.L. Spatial variation in the healthy human lung microbiome and the adapted island model of lung biogeography. *Ann Am Thorac Soc*. 2015;12(6):821-830. doi 10.1513/AnnalsATS.201501-029OC
2. Jia S., Taylor-Cousar J.L. Cystic fibrosis modulator therapies. *Annu Rev Med.* 2023;74:413-426. doi 10.1146/annurev-med-042921-021447
3. Jermiin L.S., Catullo R.A., Holland B.R. A new phylogenetic protocol: dealing with model misspecification and confirmation bias in molecular phylogenetics. *NAR Genom Bioinf.* 2020;2(2):lqaa041. doi 10.1093/nargab/lqaa041
4. Segata N., Izard J., Waldron L., Gevers D., Miropolsky L., Garrett W.S., Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12(6):R60. doi 10.1186/gb-2011-12-6-r60
5. Воронина О.Л., Рыжова Н.Н., Кунда М.С., Аксенова Е.И., Зигангирова Н.А. и др. Pseudomonas aeruginosa. Ассистенты и конкуренты в микробиоме инфицированных легких больных

муковисцидозом. *Медицинский вестник Северного Кавказа*. 2020;15(2):186-191. doi 10.14300/mnnc.2020.15045
[Voronina O.L., Ryzhova N.N., Kunda M.S., Aksenova E.I., Zigangirova N.A. et al. Pseudomonas aeruginosa. Assistants and competitors in the microbiome of infected CF patients' lungs. *Meditsinskiy vestnik Severnogo Kavkaza*. 2020;15(2):186-191. doi 10.14300/mnnc.2020.15045 (in Russian)]
6. Natalini J.G., Singh S., Segal L.N. The dynamic lung microbiome in health and disease. *Nat Rev Microbiol.* 2023;21:222-235. doi 10.1038/s41579-022-00821-x

doi 10.18699/bgrs2024-5.2-04

Реконструкция древних геномов и внутривидовая филогения *Aggregatibacter aphrophilus*

Иванова А.О.[1]*, Павлова А.В.[1], Боцманов Е.И.[1], Цедилина Т.Р.[1], Перевозчикова А.А.[2], Березина Н.Я.[2], Крицкий А.А.[1], Климук Е.И.[1], Северинов К.В.[1]

[1] *ООО «Биотехнологический камус», Москва, Россия*

[2] *Научно-исследовательский институт и Музей антропологии им. Д.Н. Анучина, Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

** aivanova@biotc.ru*

Ключевые слова: палеогеномика; филогенетика; биоинформатика; сборка *de novo*; *Aggregatibacter aphrophilus*

Мотивация и цель: *Aggregatibacter aphrophilus* – грамотрицательная, неподвижная, факультативно анаэробная палочка. В норме – комменсал ротовой полости, оппортунистический патоген, способный вызывать бактериемию, раневые инфекции, абсцессы мозга, инфекционный эндокардит у человека [1]. Входит в группу этиологических агентов инфекционного эндокардита HACEK [2]. Известно 23 полных генома штаммов данного вида, выделенных в течение 1932 – 2022 гг. В ходе палеогеномных исследований человеческих останков нами определено два древних генома *A. aphrophilus* возрастом 2170 ± 23 и 6012 ± 28 лет и проведен их функциональный и филогенетический анализ.

Методы и алгоритмы: Биоинформатический анализ проводили с использованием программ “FastP”, “SPAdes”, “metabat”, “PhyloPhlAn”, “checkm”, “roary”, “prokka”, “Blast”. Аутентификация древней ДНК осуществлялась при помощи “MapDamage2”. Филогенетический анализ был проведен с использованием “iqtree”. Для визуализации результатов использовались пакет “seaborn” (python) и “itol” [3].

Результаты: Реконструированы и проанализированы два полных генома *A. aphrophilus* возрастом 2170 ± 23 лет (AAA18C_KGZ_143BC) из зуба мужчины, погребенного в могильнике Кен-Су (Киргизия, Нарынская область), и 6012 ± 28 лет (AAA25C_RU_3899) из зуба мужчины, погребенного в могильнике Прогресс 2 (Россия, Ставропольский край).

Оба древних генома не принадлежат ни одной из филогенетических клад современных штаммов *A. aphrophilus*. В геноме AAA18C_KGZ_143BC отсутствуют регуляторные и структурные гены пилей адгезии *rcpABC/tadA-Z*. Эти гены присутствуют у всех современных клинических изолятов [4]. В геноме AAA25C_RU_3899 присутствуют некоторые гены кластера пилей адгезии (*rpcA*, *rpcB*, *rpcC*, *tadD*, *tadB*, *tadZ*). В обоих древних геномах отсутствуют гены адгезинов *aae* и *emaA*, характерные для современных клинических штаммов. Однако оба древних генома содержат кластер *lptA-G*, кодирующий основной фактор патогенности грамотрицательных бактерий, а также гены *lolA* и *oppA*, кодирующие белки, участвующие в колонизации макроорганизма. Гены, ассоциированные с устойчивостью к эльфамицину у современных изолятов, отсутствуют в древних геномах.

Выводы: Собраны и проанализированы два древних полных генома *A. aphrophilus* и продемонстрированы отличия от современных геномов этого вида, подтверждающие приобретение факторов патогенности и антибиотико-резистентности в ходе эволюции.
Финансирование: Исследование поддержано ООО «Биотехнологический кампус».

Reconstruction of ancient *Aggregatibacter aphrophilus* genomes and establishment of their intraspecific phylogeny

Ivanova A.O.[1]*, Pavlova A.V.[1], Botsmanov E.I.[1], Tsedilina T.R.[1],
Perevozchikova A.A.[2], Berezina N.Y.[2], KritskyA.A.[1], Klimuk E.I.[1], Severinov K.V.[1]
[1] *LLC "Biotechnological campus", Moscow, Russia*
[2] *Research Institute and Museum of Anthropology named after D.N. Anuchin, Lomonosov Moscow State University, Moscow, Russia*
* *aivanova@biotc.ru*

Key words: paleogenomics; phylogenetics; bioinformatics; *de novo* assembly; *Aggregatibacter aphrophilus*

Motivation and Aim: *Aggregatibacter aphrophilus* is a Gram-negative, nonmotile, facultatively anaerobic rod bacterium, a member of the HACEK group of etiological agents of infective endocarditis [1]. A part of normal oral flora [2], it is an opportunistic pathogen causing bacteremia, wound infections, brain abscesses, and infective endocarditis in humans. The genomes of 23 *A. aphrophilus* strains collected between 1932–2022 are known. The purpose of this work was to establish evolutionary relationship between modern and ancient *A. aphrophilus* isolates.
Methods and Algorithms: Bioinformatics analysis was carried out using the software "FastP", "SPAdes", "metabat", "PhyloPhlAn", "checkm", "roary", "prokka", "Blast". Authentication of ancient DNA was conducted with "MapDamage2" package. Phylogenetic analysis was performed using "iqtree". The "seaborn" python-based package and "itol" were used for the result visualization [3].
Results: Two complete genomes of *A. aphrophilus*, a 2170 ± 23 years old AAA18C_KGZ_143BC and a 6012 ± 28 years old (AAA25C_RU_3899), were reconstructed and analyzed. The ancient DNA used to reconstruct AAA18C_KGZ_143BC was isolated from a tooth of a man buried in the Ken-Su burial ground (Kyrgyzstan, Naryn region). The AAA25C_RU_3899 genome was reconstructed from DNA isolated from a tooth of a man buried in the Progress 2 burial ground (Russia, Stavropol Krai).
The ancient genomes do not belong to phylogenetic clades of modern *A. aphrophilus* strains. The AAA18C_KGZ_143BC genome lacks the *rcpABC/tadA-Z* regulatory and structural genes of adhesion pili. These genes are present in all extant clinical isolates [4]. In the AAA25C_RU_3899 genome, some adhesion pili genes (*rpcA*, *rpcB*, *rpcC*, *tadD*, *tadB*, *tadZ*) are present. In both ancient genomes, the *aae* and *emaA* adhesins genes characteristic of modern clinical strains are absent. However, both ancient genomes contain the *lptA-G* cluster encoding the main pathogenicity factor of Gram-negative bacteria, as well as the *lolA* and *oppA* genes encoding proteins involved in host

colonization. Genes associated with resistance to elfamycin in modern isolates are absent from ancient genomes

Conclusion: We have assembled and analyzed two ancient *A. aphrophilus* complete genomes. Comparisons with modern genomes of this species is consistent with acquisition of pathogenicity and antibiotic resistance factors during recent evolution of the species.

Funding: The study is supported by LLC “Biotechnological campus”.

Список литературы/References

1. Nørskov-Lauritsen N. Classification, identification, and clinical significance of Haemophilus and Aggregatibacter species with host specificity for humans. *Clin Microbiol Rev*. 2014;27(2):214-240. doi 10.1128/CMR.00103-13
2. Khaledi M., Sameni F., Afkhami H., Hemmati J. et al. Infective endocarditis by HACEK: a review. *J Cardiothorac Surg*. 2022;17(1):185. doi 10.1186/s13019-022-01932-5
3. Ancient DNA. Methods and protocols / ed. Beth Shapiro et al. Springer protocols. New York: Humana Press, 2019
4. Kittichotirat W., Bumgarner R.E., Asikainen S., Chen C. Identification of the pangenome and its components in 14 distinct Aggregatibacter actinomycetemcomitans strains by comparative genomic analysis. *PLoS One*. 2011;6(7):e22420. doi 10.1371/journal.pone.0022420

doi 10.18699/bgrs2024-5.2-05

Молекулярная филогения амеб семейства Amoebidae (Euamoebida, Tubulinea, Amoebozoa)

Камышацкая О.[1, 2]*, Смирнов А.[1]

[1] *Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*

[2] *Институт цитологии РАН, Санкт-Петербург, Россия*

* *oksana.kamyshatskaya@gmail.com*

Ключевые слова: филогения; систематика; Amoebidae; Tubulinea

Мотивация и цель: Представители семейства Amoebidae (Euamoebida, Tubulinea, Amoebozoa) являются одними из самых известных для широкой аудитории амеб. Среди них, например, *Amoeba proteus*, являющаяся популярным лабораторным объектом цитологических исследований на протяжении многих десятилетий [1–8]. Несмотря на это, группа остается парадоксальным образом недоисследованной. Основное разнообразие видов представлено в пределах родов *Chaos*, *Amoeba*, в то время как представители остальных четырех родов – *Deuteramoeba*, *Trichamoeba*, *Hydramoeba* и *Parachaos* – известны по немногочисленным, а иногда и единичным описаниям. Объем доступных молекулярных данных для семейства Amoebidae крайне ограничен. Три из шести родов остаются неизученными на молекулярном уровне, а данные по всем остальным исчерпываются всего пятью видами: двумя видами рода *Amoeba*, двумя видами рода *Chaos* и *Deuteramoeba mycophaga* (из 17 известных). Род *Polychaos*, до недавнего времени включаемый в состав семейства Amoebidae на основе морфологических признаков, был перенесен в семейство Hartmannellidae (Euamoebida, Tubulinea) на основе молекулярно-филогенетического анализа [9, 10]. Это показало несостоятельность синапоморфных признаков для групп амеб отряда Euamoebida, описывающих исключительно морфологические характеристики амеб. Таким образом, исследование эволюционных взаимосвязей между амебами отряда Euamoebida является актуальной задачей.

Дополнительную проблему при изучении этих амеб представляют сложность получения стабильных лабораторных культур, а также постоянное присутствие в их цитоплазме многочисленных пищевых вакуолей (содержащих широкий спектр про- и эукариотических организмов) и наличие прокариотических эндосимбионтов. Все это затрудняет получение неконтаминированных образцов тотальной ДНК амеб и амплификацию отдельных генов методом прямого ПЦР. Современные методы одноклеточной геномики позволяют получить молекулярные данные, используя небольшое количество клеток или отдельные ядра [9].

Методы и алгоритмы: Для изоляции ДНК клетки амеб отбирали в 40 мм пластиковые чашки Петри со стерильной фильтрованной (размер пор 0.22 μm) минеральной средой PJ [11]. В течение трех дней клетки голодали, смена минеральной среды на свежую производилась каждый день. Отсутствие видимых пищевых вакуолей проверяли, наблюдая амеб в чашках при помощи инвертированного микроскопа Leica DMI 3000, оснащенного фазово-контрастной оптикой. Далее индивидуальные клетки амеб три раза отмывали в свежей порции стерильной фильтрованной среды и помещали в 200-μl ПЦР-пробирки. Клетки

амеб с наиболее крупными ядрами были помещены в отдельные чашки со стерильной фильтрованной средой, после чего разрушены при помощи интенсивного пипетирования. Индивидуальные ядра были перенесены в свежую порцию среды при помощи тонкого капилляра. Затем отмывались последовательно в трех сменах свежей стерильной среды и, наконец, помещались в ПЦР-пробирки объемом 200 µl с небольшим объемом среды.

Экстракцию ДНК из отдельных клеток и ядер осуществляли при помощи PicoPure DNA Extraction Kit (Thermo Fischer Scientific, Уолтем, Массачусетс, США). Далее была произведена полногеномная амплификация ядерной ДНК с использованием набора для амплификации ДНК отдельных клеток REPLI-g (Qiagen, Германия) в соответствии с протоколом производителя. Полученные продукты секвенировали с помощью системы Illumina HiSeq 2500. Для каждого образца было получено 25 млн парных прочтений длиной 150 п. н. Проверка контроля качества необработанных данных последовательностей проводилась с использованием FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Для сборки митохондриального генома *de novo* использовали ассемблер SPAdes [12]. Полученные для каждого образца контиги, содержащие идентичные фрагменты гена 18S рРНК, были идентифицированы с помощью BLAST [13].

Полученные последовательности были смонтированы в выравнивании, содержащем все полученные последовательности видов амеб семейства Amoebidae, репрезентативного набора Hartmannellidae, других Tubulinea и ряда последовательностей Discosea, используемых в качестве внешней группы. Выравнивание последовательностей производилось автоматически с использованием алгоритма Muscle, реализованного в SeaView 4.0 [14]. Первоначальный выбор нуклеотидных сайтов для вывода дерева осуществляли с использованием GBlocks [15]. Филогенетический анализ проводили с использованием метода максимального правдоподобия, реализованного в программе RAxML [16], с моделью GTR + γ. Байесовский анализ того же набора данных был выполнен с использованием MrBayes 3.2.6, модели GTR с гамма-коррекцией межсайтовых вариаций скорости (8 категорий) и коварионной модели [17]. Программы RaxML и MrBayes запускались на веб-сайте Cipres V.3.3 [18].

Результаты: Главным результатом стала наиболее полная по сравнению с предыдущими работами реконструкция молекулярной филогении семейства Amoebidae. Наши данные показывают, что род *Deuteramoeba* является независимым родом семейства Amoebidae, что подтверждает морфологические данные. Он надежно группируется с родами *Chaos* и *Amoeba*. Полностью поддержанным является семейство Amoebidae в целом. Амебы рода *Polychaos* надежно группируются с амебами семейства Hartmannellidae, а не Amoebidae.

Выводы: Таким образом, семейство Amoebidae, одно из классических семейств голых лобозных амеб, является парафилетическим таксоном и требует ревизии. Таксономические критерии для этого семейства амеб требуют серьезного пересмотра. В настоящее время главной задачей остается поиск новых таксономических признаков, которые можно использовать в качестве синапоморфных для классификации этих организмов.

Финансирование: Работа выполнена при финансовой поддержке гранта РНФ № 23-24-00264 с использованием оборудования РЦ «Развитие молекулярных и клеточных технологий», «Биобанк» и «Культивирование микроорганизмов» Научного парка СПбГУ.

Molecular phylogeny of amoeba familly Amoebidae (Euamoebida, Tubulinea, Amoebozoa)

Kamyshatskaya O.[1, 2]*, Smirnov A.[1]

[1] *Saint Petersburg University, St. Petersburg, Russia*

[2] *Institute of Cytology, RAS, St. Petersburg, Russia*

* *oksana.kamyshatskaya@gmail.com*

Key words: phylogeny; systematics; Amoebidae; Tubulinea

Motivation and Aim: Amoebae of the family Amoebidae (Euamoebida, Tubulinea, Amoebozoa) are among the most popular and recognizable protists. *Amoeba proteus*, for example, has been one of the most popular objects in the multiple cytological investigations for many years [1–8]. However, the species diversity of this group remains poorly studied. The main diversity of this group is represented within the genera *Chaos* and *Amoeba*. While representatives of the other four genera *Deuteramoeba*, *Trichamoeba*, *Hydramoeba* and *Parachaos* are known from a few descriptions. Molecular data are available for a limited number of species: two species of the genus *Amoeba*, two species of the genus *Chaos* and *Deuteramoeba mycophaga* (out of 17 known). Recent study of molecular phylogeny of the genus *Polychaos* [9, 10] resulted to its transfer from family Amoebidae to family Hartmannellidae (Euamoebida, Tubulinea). This emphasized the inconsistency of synapomorphic characters, which are described exclusively by the morphological features of amoebae, for groups of amoebae within order Euamoebida. Thus, the study of the evolutionary relationships between amoebae of the order Euamoebida is an extremely urgent task.

It is really hard to obtain stable laboratory cultures of these amoebae. Many of them are large polyphagous organisms, feeding on a wide spectrum of prokaryotic and eukaryotic food. The problem is further complicated because these amoebae contain numerous endocytobionts in the cytoplasm. As a result, preparations of the total DNA obtained from these organisms usually are heavily contaminated and cannot be used for the amplification of amoebae genes by conventional PCR methods. However the modern techniques of single-cell genomics make it possible to obtain molecular data using a small number of cells or individual nuclei [9].

Methods and Algorithms: The DNA isolation was performed with individual amoeba cells or nuclei were transferred into 40 mm plastic Petri dishes filled with Millipore-filtered (0.22 μm) PJ medium [11], and left to starve for three days. Every day the medium was replaced with the fresh one. The cells were examined using an inverted Leica DMI 3000 microscope equipped with phase-contrast optics to check for the absence of visible food vacuoles. After that, one cell was collected, washed in three subsequent changes of PJ medium, and placed with 1-2 μl of the medium in 200-μl PCR tubes. Another cell was transferred to the fresh dish filled with the sterile medium and destroyed by intensive pipetting. The nucleus of the amoeba was picked out using tapered-tip glass Pasteur pipette, washed in three subsequent changes of the same medium and placed with 1–2 μl of the medium in a 200 μl PCR tube. DNA from the single cell and from the isolated nucleus was extracted using Arcturus PicoPure DNA Extraction Kit (Thermo Fischer Scientific, Waltham, MA, USA). The extraction mixture was prepared according to the manufacturer's instructions. Then, 10 μl of the mixture

was added to the tube containing the single cell/nucleus. Further, we performed the whole genome amplification of the nuclear DNA using REPLI-g Single Cell DNA Amplification Kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol. The resulting MDA products were sequenced using Illumina HiSeq 2500. For each sample there were obtained ca. 25 mln paired reads with length 150 bp. Quality control check of raw sequence data was performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). SPAdes assembler was used for de novo mitochondrial genome assembly [12]. Contigs containing individual 18S rRNA gene fragments obtained for each sample were identified using BLAST [13]. The obtained sequences weree mounted in the alignment containing all sequences of amoebae species of the family Amoebidae, a representative set of Hartmannellidae, other Tubulinea, and a number of discosean sequences used as an outgroup. Sequences were automatically aligned using the Muscle algorithm as implemented in SeaView 4.0 [14]. Initial selection of nucleotide sites for tree inference was done using GBlocks [15]. The phylogenetic analysis was performed using maximum likelihood method as implemented in RAxML program [16] with GTR + γ model. Bayesian analysis of the same dataset was performed using MrBayes 3.2.6, GTR model with gamma correction for intersite rate variation (8 categories), and the covarion model [17]. RaxML and MrBayes programs were run at Cipres V.3.3 website [18].

Results: The main result of our work is the most complete reconstruction of the molecular phylogeny of the family Amoebidae compared to the previous works. Our data indicate that the genus *Deuteramoeba* is an independent genus of the family Amoebidae, which is supported by morphological data. It is reliably grouped with the genera *Chaos* and *Amoeba*. The family Amoebidae as a whole is fully supported. Amoebas of the genus *Polychaos* are reliably grouped with amoebas of the family Hartmannellidae, not Amoebidae.

Conclusion: Thus, the family Amoebidae – one of the classic families of naked lobose amoebae, is a paraphyletic taxon and requires revision.

The taxonomic criteria for this family require serious revision. Currently, we need the new view aimed to define novel taxonomic characters that can be used as synapomorphic to define higher-level groups of Euamoebida and related taxa.

Funding: The study was supported by the Russian Science Foundation project 23-24-00264. This study utilized equipment of the Core Facility Centers “Biobank”, “Development of Molecular and Cell Technologies” and “Culture Collection of Microorganisms” of the Research Park of Saint Petersburg State University.

Список литературы/References

1. Prescott D.M. Relations between cell growth and cell division: I. Reduced weight, cell volume, protein content, and nuclear volume of *Amoeba proteus* from division to division. *Exp Cell Res*. 1955;9(2):328-337
2. Roth L.E., Obetz S.W., Daniels E.W. Electron microscopic studies of mitosis in amebae: I. *Amoeba proteus*. *J Cell Biol*. 1960;8(1):207-220
3. Jeon K.W., Jeon M.S. Cytoplasmic filaments and cellular wound healing in *Amoeba proteus*. *J Cell Biol*. 1975;67(1):243-249
4. Dembo M. Mechanics and control of the cytoskeleton in *Amoeba proteus*. *Biophys J*. 1989;55(6):1053-1080
5. Korohoda W., Golda J., Sroka J., Wojnarowicz A., Jochym P., Madeja Z. Chemotaxis of *Amoeba proteus* in the developing pH gradient within a pocket-like chamber studied with the computer assisted method. *Cell Motil Cytoskeletol*. 1997;38(1):38-53
6. Pomorski P., Krzemiński P., Wasik A., Wierzbicka K., Barańska J., Kłopocka W. Actin dynamics in *Amoeba proteus* motility. *Protoplasma*. 2007;231:31-41

7. Berdieva M., Bogolyubov D., Podlipaeva Y., Goodkov A. Nucleus-associated actin in *Amoeba proteus*. *Eur J Protistol.* 2016;56:191-199
8. Carrasco-Pujante J., Bringas C., Malaina I. et al. Associative conditioning is a robust systemic behavior in unicellular organisms: an interspecies comparison. *Front Microbiol*. 2021;12:707086
9. Kamyshatskaya O.G., Bondarenko N.I., Mesentsev Y.S., Chistyakova L.V., Nassonova E.S., Smirnov A.V. Molecular phylogeny of *Polychaos annulatum* (Amoebozoa, Tubulinea, Euamoebida) shows that genus *Polychaos* belongs to the family Hartmannellidae. *J Eukaryot Microbiol*. 2020;67:321-326
10. Kamyshatskaya O., Bondarenko N., Nassonova E., Smirnov A. *Polychaos centronucleolus* n. sp. – a new terrestrial species of the genus *Polychaos* (Amoebozoa, Tubulinea) with nontypical nuclear structure. *Eur J Protistol*. 2021;77:125759
11. Prescott D.M., James T.W. Culturing of *Amoeba proteus* on *Tetrahymena*. *Exp Cell Res*. 1955;81:256-258. doi 10.1016/0014-4827(55)90067-7
12. Bankevich A., Nurk S., Antipov D. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol*. 2012;19:455-477
13. Altschul S.F., Madden T.L., Schaffer A.A. et al. Gapped BLAST and PSIBLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389-3402
14. Gouy M., Guindon S., Gascuel O. Sea view version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 2010;27:221-224
15. Cestresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Ecular Biol Evol* 2000;17(4):540-552
16. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312-1313
17. Ronquist F., Huelsenbeck J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572-1574
18. Miller M.A., Pfeiffer W., Schwartz T. Creating the CIPRES science gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE*)*. New Orleans (USA), 2010

doi 10.18699/bgrs2024-5.2-06

Выявление участков пластомов, потенциально пригодных для метабаркодинга видов растений сем. Poaceae

Криницына А.[1, 2]*, Никитина О.[1, 2], Широбоков В.[1], Никитин П.[1], Северова Е.[1]

[1] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

[2] *ООО «Эвоген», Москва, Россия*

* *ankrina@gmail.com*

Ключевые слова: Poaceae; идентификация пыльцы; метабаркодинг; пластомы

Мотивация и цель: Аллергены, содержащиеся в переносимой по воздуху пыльце, вызывают астму и аллергический ринит, от которых страдают около 300 миллионов человек по всему миру. Одним из наиболее распространенных аллергенов является пыльца злаковых (Poaceae). Это семейство включает более 11 000 видов [1], большинство из которых опыляются ветром. Пыльца злаков присутствует в воздухе в течение очень длительного периода, поэтому пациенты с аллергией на пыльцу злаковых могут страдать от симптомов сезонного аллергического ринита долгое время. Пыльца всех видов злаков относится к единому палиноморфологическому типу и при проведении аэробиологического мониторинга стандартными методами идентифицируется только до уровня семейства. В то же время аллергенность пыльцы разных видов злаков различается, поэтому возникает необходимость детализации кривых пыления. Идентификация биоразнообразия с помощью метабаркодирования представляет собой еще один вариант идентификации состава сложных смесей, в том числе для идентификации и количественного определения пыльцы и образцов из окружающей среды. Наиболее популярными генными маркерами, используемыми для филогенетической классификации, являются внутренние транскрибируемые спейсеры (ITS1 и ITS2), а также участки пластомной ДНК – часть последовательностей генов *rbcL* и *matK*. Технология секвенирования длинных участков позволяет расширить спектр маркерных последовательностей, перспективных для определения систематической принадлежности растений в сложных растительных смесях. Цель нашей работы – выявить участки пластомной ДНК, пригодные для определения видовой принадлежности пыльцы злаков в аэробиологических образцах.

Методы и алгоритмы: На основании последовательностей полных пластомов 52 видов, относящихся к сем. Poaceae, был проведен поиск наиболее вариативных участков с дальнейшей теоретической оценкой возможности их использования для метабаркодинга. Множественное выравнивание последовательностей пластомов проводили с помощью программы MAFFT v7.511. Поиск наиболее полиморфных участков выполняли следующим образом: с помощью Jupyter notebook строили гистограмму плотности полиморфизмов по сделанному выравниванию с использованием python 3.9.12 и библиотек AlignIO и Seaborn. Алгоритмом разметки регионов по плотности гэпов в выравнивании с шагом в 5000 и 1000 для лучшего масштабирования. С использованием выявленных участков проводили кластеризацию видов в программе MEGA v 11.0.13 (maximum likelihood (ML), bootstrap = 1000). Аннотацию выбранных участков выполняли с использованием MPI-MP CHLOROBOX – GeSeq (GeSeq – annotation of Organellar Genomes). Подбор праймеров на выявленные участки пластомов осуществляли в программе Primer-Blast. Получение продуктов амплификации проводили с

использованием в качестве матрицы ДНК, полученной из искусственных смесей пыльцы разных видов злаков, протокол ПЦР описан в работе [2].
Результаты: В результате проведенного анализа нами были отобраны семь наиболее вариативных фрагментов участков пластомной ДНК размером 800–1000 п. н. Из них только четыре позволили провести кластеризацию одноименных родов в общие группы. Филогенетические отношения видов отражают деревья, полученные с использованием участков r2_cpDNA_3 (фрагмент гена *rps16*) и r2_cpDNA_4 (фрагмент пластома, содержащий участок генов *rps16* и *trnQ*-UUG). Участки пластома, содержащие ген *rps16*, а также межгенные спейсеры с соседними генами, показали существенную значимость в филогенетических исследованиях разных подсемейств злаковых (например, [3, 4]. Менее достоверный результат был получен с использованием участков r7_cpDNA_1-2 (гены *trnG*-UCC, *trnT*-GGU и *trnE*-UUC). Вариабельные участки r22_cpDNA_1 (*ndhF*) и r22_cpDNA_2 (*rpl32*) оказались непригодными для кластеризации одноименных родов в общие группы. Вероятно, участок пластомного генома, содержащий гены *ndh*, отражает экологическую адаптацию, позволяющую покрытосеменным выживать во многих стрессовых условиях, поддерживая эффективный фотосинтез (Sabater, 2021). Для выбранных четырех вариантов последовательностей удалось получить продукт амплификации указанного размера с использованием ДНК из искусственных пыльцевых смесей.
Выводы: Для определения систематической принадлежности возможно использовать фрагменты пластома, содержащие фрагмент генов *rps16* и *trnQ*-UUG. Дополнительно можно использовать последовательности генов и межгенных спейсеров *trnG*-UCC, *trnT*-GGU и *trnE*-UUC.
Финансирование: Исследование выполнено в рамках госзадания Московского государственного университета им. М.В. Ломоносова № 121032500082-2.

Identification of regions of chloroplast DNA potentially useful for metabarcoding of plant species of the fam. Poaceae

Krinitsina A.[1, 2]*, Nikitina O.[1, 2], Shirobokov V.[1], Nikitin P.[1], Severova E.[1]
[1] *Lomonosov Moscow State University, Moscow, Russia*
[2] *LCT Evogen, Moscow, Russia*
* *ankrina@gmail.com*

Key words: Poaceae; pollen identification; metabarcoding; plastome

Motivation and Aim: Allergens in airborne pollen cause asthma and allergic rhinitis, which affect about 300 million people worldwide. One of the most common allergens is pollen from cereals (Poaceae). This family includes more than 11,000 species [1], most of which are wind pollinated. Grass pollen is present in the air for a very long period, so patients allergic to grass pollen may suffer from symptoms of seasonal allergic rhinitis for a long period of time. The pollen of all Poaceae belongs to a single palynomorphologic type and is identified only to the family level during aerobiologic monitoring by standard methods. At the same time, allergenicity of pollen of different species differs, therefore, there is a need to detail pollen curves. Biodiversity identification using metabarcoding is another option for identifying the composition of complex mixes, including for the identification and quantification of pollen and environmental samples. The most popular gene markers used for molecular phylogeny are internally transcribed spacers (ITS 1 and ITS2), as well as regions of chloroplast DNA – part of

the *rbcL* and *matK* gene. The technology of sequencing long DNA regions allows us to expand the range of marker sequences promising for determining the identity of plants in complex mixtures. The aim of our work was to identify regions of chloroplast DNA useful for the determination the species identity of cereal pollen in aerobic samples.

Methods and Algorithms: Based on the sequences of complete cpDNA of 52 species of Poaceae, the most variable sites were searched and further theoretically evaluated for their use for metabarcoding. Multiple sequence alignment of plastome sequences was performed using the MAFFT v7.511. The search for the most polymorphic regions was performed as follows: using Jupyter notebook we constructed a histogram of polymorphism density based on the made alignment using python 3.9.12 and AlignIO and Seaborn libraries. Algorithm partitioned regions by density gaps in the alignment with steps of 5000 and 1000 for better scaling. Using the identified regions, species clustering was performed using MEGA v 11.0.13 (maximum likelihood (ML), bootstrap = 1000). Annotation of selected regions was performed using MPI-MP CHLOROBOX – GeSeq (GeSeq – annotation of Organellar Genomes). Primers for the identified plastome sites was performed using Primer-Blast. Amplification products were obtained using DNA obtained from artificial mixes of pollen of different cereal species as a matrix; the PCR protocol is described in [2].

Results: As a result of our analysis, we selected 7 most variable fragments of 800–1000 bp cpDNA regions. Only 4 of them allowed clustering of the same-named genera into common groups. Phylogenetic relationships of species are reflected by phylogenetic trees obtained using the r2_cpDNA_3 (part of *rps16* gene) and r2_cpDNA_4 (cpDNA regions containing the part of *rps16* and *trnQ*-UUG genes). cpDNA regions containing the *rps16* gene as well as intergenic spacers with neighboring genes have shown significant significance in phylogenetic studies of different subfamilies (e.g. [3, 4]). Less reliable results were obtained using the r7_cpDNA_1-2 (*trnG*-UCC, *trnT*-GGU and trnE-UUC genes). Variable sites of r22_cpDNA_1 (*ndhF*) and r22_cpDNA_2 (*rpl32*) were not suitable for clustering the same-named genera into common groups. It is likely that the region of the cpDNA containing the *ndh* genes reflects an ecological adaptation that allows covered seeds to survive under many stress conditions by maintaining efficient photosynthesis (Sabater, 2021). For the selected 4 sequence variants, we were able to obtain an amplification product of the indicated size using DNA from artificial pollen mixtures.

Conclusion: Use fragments of cpDNA containing a part of the *rps16* and *trnQ*-UUG genes is probably a good fit for identification of the pollen species. It is also possible to additionally use sequences of the genes and intergenic spacers trnG-UCC, trnT-GGU, and trnE-UUC.

Funding: The study is supported to the scientific programs of M.V. Lomonosov Moscow State University No. 121032500082-2.

Список литературы/References

1. Christenhusz M.J.M., Byng J.W. The number of known plants species in the world and its annual increase. *Phytotaxa*. 2016;261(3):201-217
2. Krinitsina A.A. et al. Aerobiological monitoring and metabarcoding of grass pollen. *Plants*. 2023;12(12):2351-2023
3. Guo-Ye G. et al. Phylogenetic relationships among Leymus and related diploid genera (Triticeae: Poaceae) based on chloroplast trnQ–rps16 sequences. *Nor J Bot.* 2014;32(5):658-666
4. Krawczyk K. et al. Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: *Stipa*). *Sci Rep.* 2018;8:1924

doi 10.18699/bgrs2024-5.2-07

Метавиромное исследование мазков из ротовой полости и фекалий ежей, обитающих на территории Российской Федерации

Лукина-Гронская А.В.[1]*, Корнеенко Е.В.[1], Чудинов И.К.[1], Машкова С.Д.[1], Сонец И.В.[1], Семашко Т.А.[1], Литвинова Е.М.[2], Феоктистова Н.Ю.[3], Синькова М.А.[4], Сперанская А.С.[1]

[1] *Научно-исследовательский институт системной биологии и медицины Роспотребнадзора, Москва, Россия*

[2] *Московский государственный университет им. М.В. Ломоносова, биологический факультет, Москва, Россия*

[3] *Институт проблем экологии и эволюции им. А.Н. Северцова РАН, Москва, Россия*

[4] *Зоологический музей Московского государственного университета им. М.В. Ломоносова, Москва, Россия*

** lukina.al98@gmail.com*

Ключевые слова: ежи; *E. roumanicus*; вирусное разнообразие; метавиром; вирусы позвоночных; коронавирусы; бетакоронавирусы; EriCoV; *Betacoronavirus Erinaceus;* маммарена-вирусы; MEMV

Мотивация и цель: В ряде зарубежных исследований показано, что в ежах, обитающих на территории Европы, найдены вирусы, которые потенциально опасны для человека и других млекопитающих. Исследований вирома ежей на данный момент не проводилось: большинство работ ограничиваются описанием единичных случаев. Известно, что ежи являются носителями клещевого энцефалита (сем. Flaviviridae) [1], вируса бешенства (сем. Rhabdoviridae) [2], герпесвирусов (сем. Herpesviridae) и парамиксовирусов (сем. Paramyxoviridae, англ. Belerina virus) [3]. В 2014 г. в Германии впервые был обнаружен бетакорона-вирус у *Erinaceus europaus* [4]. Впоследствии бетакоронавирусы ежей были также выявлены во Франции [5], Италии [6], Великобритании [7], Польше [8] и выделены в отдельную кладу EriCoV, филогенетически близкую к MERS-CoV. Другой бетакоронавирус HKU31 был выявлен у *E. amurensis* в Китае [9]. Также известно, что ежи являются носителями вирусов, принадлежащих к семейству Arenaviridae. В 2023 г. вирус Mecsek Mountains virus (англ. MEMV), относящийся к роду Mammarenaviruses, был обнаружен у ежей, обитающих на территории Венгрии [10]. В Европе распространены лишь два вируса из семейства аренавирусов – MEMV и вирус лимфоцитарного хориоменингита (англ. Lymphocytic choriomeningitis, LCMV), последний из которых переносится грызунами. Секвенирование вирома ежей, обитающих на территории Российской Федерации, не проводилось. Основной целью данного исследования являются характеристика вирусного разнообразия у *E. roumanicus*, пойманных на территории России, с использованием методов высокопроизводительного секвенирования, а также характеристика потенциально зоонозных вирусов.

Методы и алгоритмы: В исследование было включено 21 животное вида *Erinaceus roumanicus*, пойманных на территории европейской части России, а

также на территории Сибири (Академгородок, Новосибирск). Большая часть животных была поймана профессиональными зоологами, остальные животные были предоставлены сотрудниками реабилитационного центра Зоологического музея МГУ. У каждого животного были взяты или только ротоглоточные, или ротоглоточные и ректальные мазки. Высокопроизводительное секвенирование тотальной РНК было произведено на платформе BGI DNBSEQ-G400. Полученные данные были дополнительно подтверждены сиквенсом на платформе Oxford Nanopore Technologies. Парные прочтения обрабатывались с использованием fastp v.0.23.4 и более поздних версий Kraken2 версии 2.1.3 с базой данных pluspf (включая прокариоты, эукариоты и вирусы) и параметром -confidence 0.5. Результаты анализа представленности вирусных семейств были использованы для визуализации составных гистограмм, нормализованных по сумме подсчетов прочтений каждого вирусного семейства/рода, присутствующего в образце. Сборка геномов осуществлялась с помощью SPAdes в режиме работы --meta, таксономическая классификация контигов проводилась с помощью kraken2 и базы pluspf. Детализированная таксономическая классификация осуществлялась при помощи базы ICTV (International Committee on Taxonomy of Viruses). Множественное выравнивание было сделано при помощи программы Mafft (версия 7.505). Филогенетические деревья максимального правдоподобия (ML) были построены с использованием IQ-TREE (версия 2.0.7).

Результаты: Всего было отсеквенировано 34 ротоглоточных и ректальных мазка, полученных от 21 ежа. Образец считался положительным, если более 20 прочтений были классифицированы как принадлежащие к одному семейству вирусов. В результате было идентифицировано 26 семейств вирусов, которые включали вирусы растений, насекомых, грибов и бактериофагов. Было выявлено всего 6 семейств вирусов, ассоциированных с вирусами, переносчиками которых являются позвоночными. В силу того, что лаборатория работает с клиническими образцами SARS-CoV-2, образцы дополнительно проверялись альтернативными биофинформатическими методами. В результате 9 из 10 образцов были оценены как ложноположительные. Еж с идентификатором 22_15(MOS) оказался переносчиком двух разных коронавирусов и двух разных аренавирусов. Оба штамма коронавируса имеют одинаковый размер. Филогенетический анализ полных геномов бетакоронавирусов показал, что оба полученных штамма принадлежат к подроду Merbecovirus и относятся к той же кладе EriCoV, что и бетакоронавирусы, обнаруженные в Европе. Организация генома двух штаммов коронавируса была аналогична организации генома других представителей бетакоронавирусов. Попарное выравнивание показывает, что оба генома имеют 98.87 % идентичности и различаются по 342 нуклеотидам. В четырех образцах ежей (22_6(BEL), 22_7(MOS), 22_15(MOS), 23_1(MOW)) были обнаружены прочтения, классифицированные как принадлежащие к семейству Arenaviridae, роду Mammarenavirus. В образце 22_15(MOS) обнаружены два штамма аренавирусов. Длина L-сегмента генома первого штамма составляет 7331 п. о. и имеет 73.55 % идентичности последовательности с L-сегментом маммаренавируса MEMV (номер GenBank OP191655.1). Длина L-сегмента второго штамма составляет 7515 п. о. и имеет 74.3 % идентичности последовательности с L-сегментом маммаренавируса Alxa virus (номер GenBank NC_078105.1).

Выводы: Обитающие на территории России ежи вида *E. roumanicus* являются носителями бетакоронавирусов, которые относятся к той же кладе EriCoV, что и

коронавирусы, обнаруженные ранее на территории европейских стран. Данный вид ежей является носителем аренавирусов, близких к найденным ранее в ежах того же вида вирусов в Венгрии.
Финансирование: Работа выполнена на базе НИИ СБМ Роспотребнадзора в рамках госзадания «Разработка алгоритмов для выявления новых, уникальных последовательностей ДНК или РНК в метагеномах и их фенотипическая характеристика in vitro», номер 12203090069-4.

Metavirome research of oral swabs and feces of hedgehogs from Russian Federation

Lukina-Gronskaya A.V.[1], Korneenko E.V.[1], Chudinov I.K.[1], Mashkova S.D.[1], Sonets I.V.[1], Semashko T.A.[1], Litvinova E.M.[2], Feoktistova N.Y.[3], Sinkova M.A.[4], Speranskaya A.S.[1]

[1] *Scientific Research Institute for Systems Biology and Medicine, Federal Service on Consumers' Rights Protection and Human Well-Being Surveillance, Moscow, Russia*

[2] *Lomonosov Moscow State University, Biological Department, Moscow, Russia*

[3] *A.N. Severtsov Institute of Ecology and Evolution, RAS, Moscow, Russia*

[4] *Zoological Museum of Lomonosov Moscow State University, Moscow, Russia*

* *lukina.al98@gmail.com*

Key words: hedgehogs; *E. roumanicus*; viral diversity; metavirome; vertebrae-associated viruses; coronavirus; betacoronaviruses; EriCoV; *Betacoronavirus Erinaceus*; mammarenaviruses; MEMV

Motivation and Aim: Some foreign studies have shown that European hedgehogs contain many viruses that are potentially dangerous to humans and other mammals. There is no full research of hedgehogs` virome: most studies of viral findings are limited to single cases description. It is known that hedgehogs are carriers of tick-borne encephalitis (family Flaviviridae) [1], rabies virus (family Rhabdoviridae) [2], herpes viruses (family Herpesviridae) [3] and paramyxoviruses (family Paramyxoviridae, en. Belerina virus) [4]. In 2014, betacoronavirus was first discovered in *Erinaceus europaus* in Germany [4]. Subsequently, betacoronaviruses were also discovered in France [5], Italy [6], Great Britain [7], Poland [8] and were separated to an isolated clade EriCoV, phylogenetically close to MERS-CoV. Another related betacoronavirus, HKU31, was identified in *E. amurensis* in China [9]. Hedgehogs are also known as carries of viruses belonging to the Arenaviridae family. In 2023, the Mecsek Mountains virus (MEMV), belonging to the genus Mammarenaviruses, was identified in hedgehogs in Hungary [10]. Only two viruses from this genus are known to be widespread in Europe – MEMV and Lymphocytic choriomeningitis (LCMV), the last one is transmitted by rodents. At the moment, little is known about the virome of hedgehogs. Currently, there is no information on sequencing the viral genomes of hedgehogs living in the Russian Federation. The main goal of this study is to characterize viral diversity in hedgehogs collected in Russia using high-throughput sequencing methods and characterize potentially zoonotic viruses.
Methods and Algorithms: The study included 21 animals of the species *Erinaceus roumanicus*, caught in the European part of Russia, as well as in Siberia

(Akademgorodok, Novosibirsk). Most of the animals were caught by professional zoologists, the rest of the animals were provided by employees of the rehabilitation center of the Zoological Museum of Moscow State University. Only oropharyngeal or oropharyngeal and rectal swabs were collected from each animal. High-throughput total RNA sequencing was performed on the BGI DNBSEQ-G400 platform. The data obtained were further confirmed by sequencing on the Oxford Nanopore Technologies platform. Paired-end reads were processed using fastp v.0.23.4 and later versions of Kraken2 v.2.1.3 with pluspf database (including prokaryotes, eukaryotes and viruses) and -confidence parameter 0.5. Viral abundance results were used to visualize composite histograms normalized by the sum of read counts for each viral family/genus present in the sample. Genome assembly was carried out using SPAdes in the --meta operating mode, taxonomic classification of contigs was carried out using kraken2 and the pluspf database. Detailed taxonomic classification was carried out using the ICTV (International Committee on Taxonomy of Viruses) database. The complete genomes including newly discovered coronaviruses were aligned using Mafft v7.505 (2022/Apr/10). The Maximum likelihood (ML) phylogenetic trees were constructed using IQ-TREE (version 2.0.7).

Results: A total of 34 oropharyngeal and rectal swabs obtained from 21 hedgehogs were sequenced. A sample was considered positive if more than 20 reads were classified as belonging to the same virus family. As a result, 26 families of viruses were identified, which included viruses of plants, insects, oomycetes, fungi and bacteriophages. 6 families of viruses associated with viruses transmitted by vertebrates were identified. Due to the fact that the laboratory works with clinical samples of SARS-CoV-2, the samples were additionally tested using alternative bioinformatics methods. As a result, 9 out of 10 samples were considered false positive. The hedgehog with identification number 22_15 (MOS) turned out to be a carrier of two different coronaviruses and two different arenaviruses. Both coronaviruses have the same size. Phylogenetic analysis of complete betacoronavirus genomes showed that both strains belong to the subgenus Merbecovirus and belong to the same EriCoV clade as betacoronaviruses found in Europe. The genome organization of the two coronavirus strains was similar to other betacoronaviruses. Pairwise alignment shows that both genomes have 98.87 % identity and differ by 342 nucleotides. In four hedgehog samples (22_6(BEL), 22_7(MOS), 22_15(MOS), 23_1(MOW) reads were classified as belonging to the family Arenaviridae, genus Mammarenavirus. Two strains of arenaviruses were detected in sample 22_15(MOS). The length of the L-segment of the genome of the first strain is 7331 b.p. and has 73.55 % sequence identity with the L segment of Mecsek Mountains mammarenavirus (GenBank number OP191655.1). The length of the L-segment of the second strain is 7515 b.p. and has 74.3 % sequence identity with the L-segment of the Alxa virus mammarenavirus (GenBank number NC_078105.1).

Conclusion: The results of this study prove that *Erinaceus roumanicus* Russia are carriers of betacoronaviruses, which are included in the same EriCoV clade as coronaviruses previously discovered in European countries. This species of hedgehog is a carrier of arenaviruses close to viruses previously found in hedgehogs of the same type in Hungary.

Funding: This study was financially supported by the Research Institute of SBM of Rospotrebnadzor within the framework of the state task “Algorithm development for identifying new, unique DNA or RNA sequences in metagenomes and their phenotypic characteristics in vitro”, number 12203090069-4.

Список литературы/References

1. Schönbächler K., Hatt J., Silaghi C. et al Confirmation of Tick-borne encephalitis virus in an European hedgehog (Erinaceus europaeus). *Schweiz Arch Tierheilkd*. 2019;161(1):23-31. doi 10.17236/sat00191
2. Faragó Z. Rabid hedgehog in inner-city area of Budapest. *Orv Hetil*. 1997;138(36):2231-2232. doi 10.1556/650.1997.09.07
3. Hydeskov H.B., Dastjerdi A., Hopkins K.P. et al. Detection and characterisation of multiple herpesviruses in free-living Western European hedgehogs (Erinaceus europaeus). *Sci Rep*. 2018;8:13942
4. Vanmechelen B., Vergote V., Merino M., Verbeken E., Maescorresponding P. Common occurrence of Belerina virus, a novel paramyxovirus found in Belgian hedgehogs. *Sci Rep*. 2020;10(1):19341. doi 10.1038/s41598-020-76419-1
5. Monchatre-Leroy E., Boué F., Boucher J.M. et al. Identification of alpha and beta coronavirus in wildlife species in France: bats, rodents, rabbits, and hedgehogs. *Viruses*. 2017;9(12):364. doi 10.3390/v9120364
6. Delogu M., Cotti C., Lelli D. et al. Eco-virological preliminary study of potentially emerging pathogens in hedgehogs (Erinaceus Europaeus) recovered at a wildlife treatment and rehabilitation center in northern Italy. *Animals (Basel)*. 2020;10(3):407. doi 10.3390/ani10030407
7. Saldanha I.F., Lawson B., Goharriz H. et al. Extension of the known distribution of a novel clade C betacoronavirus in a wildlife host. *Epidemiol Infect*. 2019;147:e169. doi 10.1017/S0950268819000207
8. Pomorska-Mól M., Ruszkowski J.J., Gogulski M., Domanska-Blicharz K. First detection of Hedgehog coronavirus 1 in Poland. *Sci Rep*. 2022;12(1):2386. doi 10.1038/s41598-022-06432-z
9. Lau S.K.P., Luk H.K.H., Wong AC.P. et al. Identification of a novel betacoronavirus (merbecovirus) in amur hedgehogs from China. *Viruses*. 2019;11(11):980. doi 10.3390/v11110980
10. Reuter G., Boros Á., Takáts K., Mátics R., Pankovics P. A novel mammarenavirus (family Arenaviridae) in hedgehogs (*Erinaceus roumanicus*) in Europe. *Arch Virol*. 2023;168(7):174. doi 10.1007/s00705-023-05804-8

doi 10.18699/bgrs2024-5.2-08

Полногеномное секвенирование микроспоридий рода *Nosema* – внутриклеточных паразитов чешуекрылых насекомых

Малыш С.М.[1]*, Румянцева А.С.[1], Сахабеев Р.Г.[2], Ефейкин Б.Д.[3], Данилов Л.Г.[4], Токарев Ю.С.[1]

[1] *Всероссийский институт защиты растений, Санкт-Петербург, Россия*

[2] *Санкт-Петербургский государственный технологический институт (технический университет), Санкт-Петербург, Россия*

[3] *Институт проблем экологии и эволюции им. А.Н. Северцова РАН, Москва, Россия*

[4] *Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*

* *malyshsvetlana@gmail.com*

Ключевые слова: микроспоридии; эволюция генома; мультигенная филогения; мультилокусное генотипирование; генодиагностика; близкородственные виды

Мотивация и цель: Микроспоридии – родственные грибам протисты, наиболее широко распространенные в качестве паразитов членистоногих и других животных [1, 2]. В связи с адаптацией к внутриклеточному паразитизму, они характеризуются высоким уровнем редукции клеточной организации и компактизации генома и служат важной моделью исследований молекулярной и клеточной биологии. Практическое значение микроспоридий определяется способностью вызывать заболевания экономически значимых видов беспозвоночных и позвоночных, а также человека [3–5]. Описано свыше 200 родов микроспоридий, содержащих около 1300 видов, для более 50 из которых получены полногеномные сиквенсы [6]. В связи с таким маленьким количеством геномных последовательностей и все большим количеством новых изолятов возникают трудности с проведением видовой идентификации близкородственных видов микроспоридий. Цель работы – получение исходных данных и первичная оценка результатов полногеномного секвенирования микроспоридий *Nosema pyrausta* из кукурузного мотылька *Ostrinia nubilalis* и *Nosema* sp. NspHA22 из хлопковой совки *Helicoverpa armigera* в сравнении с референтными последовательностями *N. bombycis* из тутового шелкопряда *Bombyx mori* и *N. antheraeae* из китайского дубового шелкопряда *Antheraea pernyi*. Это необходимо для изучения внутривидового полиморфизма, поиска молекулярных маркеров для разграничения близкородственных видов, выявления генетических детерминант факторов вирулентности, филогеномного анализа и других исследований молекулярной генетики энтомопатогенных микроспоридий. В частности, известно, что по общепринятому диагностическому фрагменту гена рРНК вышеуказанные изоляты из тутового шелкопряда и хлопковой совки неотличимы друг от друга, однако их биологические свойства существенно различаются, что не позволяет относить их к одному виду [7].

Методы и алгоритмы: Для накопления спор использовали гусениц первого лабораторного поколения сибирского шелкопряда *Dendrolimus sibiricus*, любезно предоставленных Александром А. Агеевым (филиал ВНИИЛМ «Центр лесной пирологии», Красноярск). Гусениц второго возраста заражали в дозировке 1 млн

спор на гусеницу и содержали в чашках Петри на хвое пихты сибирской при +24 °С. Через 2 месяца гусениц вскрывали, препарировали ткани внутренних органов в дистиллированной воде и гомогенизировали вручную в стеклянном гомогенизаторе с тефлоновым пестиком, центрифугировали при 1000 g в течение 5 мин, осадок промывали в дистиллированной воде, удаляя верхние слои и повторно центрифугируя. Количество спор подсчитывали в камере Горяева. Образец, содержащий от 1 до 1.5 млрд чистых спор, растирали в фарфоровой ступке с добавлением жидкого азота и использовали для экстракции ДНК фенольно-хлороформным методом [8]. Качество геномной ДНК определяли с помощью гель-электрофореза и наноспектрофотометра Implen, а качество ридов – утилитой FastQC. Полногеномное секвенирование осуществляли на платформе Illumina Nextseq в центре коллективного пользования Сколковского института науки и технологий (Сколтех, Москва). Сборка генома осуществлялась программой MEGAHIT. Множественное выравнивание нуклеотидных последовательностей выполнялось с использованием алгоритма ClustalW, построение филогенетических деревьев осуществлялось методом максимального правдоподобия.

Результаты: Геном *N. pyrausta* депонирован в Генбанк под номером доступа GCA_001924031.1. Приблизительный размер генома составил 15.9 Mb, количество предсказанных генов 4468, что очень близко к соответствующим показателям близкородственного вида *N. bombycis.* При этом для другого вида этого рода, а именно *N. antheraeae*, размер генома примерно в четыре раза меньше (табл. 1).

Таблица 1. Основные характеристики генома трех близкородственных видов микроспоридий рода *Nosema*

Характеристика	*N. bombycis*	*N. antheraeae*	*N. pyrausta*
Номер доступа в Генбанке	NA30919	NA183977	GCA_001924031.1
Название изолята	NbCQ1	YY	Npyr2020
Размер генома, Mb	15.7	6.6	15.9
N50	57394	1883	8050
Количество генов	4458	3413	4468
GC состав, %	31	28	34.5
Количество контигов	3558	?	4115
Количество скэффолдов	1607	6215	1880
Ссылка	Pan et al., 2013	Pan et al., 2013	Собственные данные

Для генома NspHA22 на момент подачи тезисов сборка и аннотирование не завершены из-за более высокой примеси геномной ДНК хозяина, что требует дополнительной работы по фильтрации неспецифических сигналов. Тем не менее первичные полногеномные данные обоих изолятов микроспоридий использованы для извлечения нуклеотидных последовательностей отдельных генов, наиболее часто используемых для мультилокусного генотипирования и мультигенных филогенетических реконструкций: белка теплового шока HSP70 и РНК-полимеразы RPB1, а также гексокиназы, дополнительно предложенной для меж- и внутривидовой дифференциации микроспоридий [9]. Анализ соответствующих нуклеотидных последовательностей, изучаемых микроспоридий, показал их разделение на две сестринские филогенетические линии внутри рода *Nosema.* В

одну группу вошли *N. bombycis* и NspHA22, в другую – *N. pyrausta* и *N. antheraeae*. Сходство по каждому из трех генов внутри группы составило около 98–99 %, а между группами – 92–93 %. Это позволяет достаточно четко разграничить изучаемые изоляты, однако требует секвенирования соответствующих локусов для наиболее близкородственных форм, так как низкая частота нуклеотидных замен не дает возможности сконструировать праймеры, специфичные для конкретного вида (изолята).

Выводы: Считается, что такой относительно большой, по сравнению с другими видами рода *Nosema*, размер генома *N. bombycis* обусловлен недавним событием геномной дупликации [10]. Следовательно, данное событие затронуло как *N. bombycis*, так и *N. pyrausta*, а поскольку круг их восприимчивых хозяев существенно различается [11, 12], полученная информация может послужить дальнейшему поиску генетических детерминант вирулентности энтомопатогенных микроспоридий, вызывающих опасные заболевания культивируемых насекомых и требующих разработки подходов для профилактики и лечения [13], а также перспективных с точки зрения защиты растений [14]. Геномные данные послужат также для поиска вариабельных участков генома для экспресс-диагностики, в частности, для разработки видоспецифичных праймеров.

Финансирование: Исследование выполнено в рамках проекта РНФ № 23-16-00262 и проекта Госзадания лаборатории молекулярной защиты растений ФГБНУ ВИЗР № FGEU-2022-0019 (№ госрегистрации 122032900144-2).

Whole genome sequencing of microsporidia of the genus *Nosema* – intracellular parasites of lepidopteran insects

Malysh S.[1]*, Rumiantseva A.[1], Sakhabeev R.[2], Efeikin B.[3], Danilov L.[4], Tokarev Y.[1]

[1] *All-Russian Institute of Plant Protection, Saint Petersburg, Russia*

[2] *St. Petersburg State Institute of Technology, Saint Petersburg, Russia*

[3] *A.N. Severtsov Institute of Ecology and Evolution, Moscow, Russia*

[4] *Saint Petersburg State University, Saint Petersburg, Russia*

* *malyshsvetlana@gmail.com*

Key words: microsporidia; genome evolution; multigene phylogeny; multilocus sequence typing; molecular diagnostics; closely related species

Motivation and Aim: Microsporidia are protists related to fungi, most widely distributed as parasites of arthropods and other animals [1, 2]. Due to adaptation to intracellular parasitism, they are characterized by a high level of reduction of cellular organization and genome compaction and serve as an important model for research in molecular and cellular biology. The practical importance of microsporidia is determined by the ability to cause diseases of economically important species of invertebrates and vertebrates, as well as humans [3–5]. Over 200 microsporidia genera have been described, containing about 1300 species, for more than 50 of which full genome sequences have been obtained [6]. And because of such a low number of whole genome sequences and an increasing number of newly discovered isolates, difficulties began to arise in identifying the species in closely related microsporidian taxa. The aim of the work was to perform whole-genome sequencing of microsporidian isolates *Nosema pyrausta* from the European corn

borer *Ostrinia nubilalis* and *Nosema* sp. NspHA22 from the cotton bollworm *Helicoverpa armigera*. The reference sequences were those of *N. bombycis* from the silkworm *Bombyx mori* and *N. antheraeae* from the Chinese oak silkworm *Antheraea perny* to study intraspecific polymorphism, search for molecular markers to distinguish closely related species, identify genetic determinants of virulence factors, phylogenomic analysis and other studies of the molecular genetics of entomopathogenic microsporidia. In particular, it is known that according to the generally accepted diagnostic fragment of the rRNA gene, the above isolates from the silkworm and the cotton bollworm are indistinguishable from each other, but their biological properties differ significantly, which does not allow them to be classified as the same species [7].

Methods and Algorithms: To accumulate spores, we used caterpillars of the first laboratory generation of the Siberian silkworm *Dendrolimus sibiricus*, kindly provided by Alexander A. Ageev (branch of the VNIILM Center for Forest Pyrology, Krasnoyarsk). Second instar caterpillars were infected at a dosage of 1 million spores per caterpillar and kept in Petri dishes on Siberian fir needles at +24 °C. After 2 months, the caterpillars were dissected, the tissues of the internal organs were dissected in distilled water and homogenized manually in a glass homogenizer with a Teflon pestle, centrifuged at 1000 *g* for 5 min, the sediment was washed in distilled water, removing the upper layers and centrifuging again. The number of spores was counted in Goryaev's chamber. A sample containing 1 to 1.5 billion purified spores was ground in a porcelain mortar with liquid nitrogen and used for DNA extraction using the phenol-chloroform method [8]. The quality of genomic DNA was determined using gel electrophoresis and Implen nanospectrophotometer, and the quality of reads was determined using the FastQC utility. Whole-genome sequencing was carried out on the Illumina Nextseq platform at the shared use center of the Skolkovo Institute of Science and Technology (Skoltech, Moscow) using the MEGAHIT program. Multiple alignment of nucleotide sequences was performed using the ClustalW algorithm, and the phylogenetic trees were constructed using the maximum likelihood method.

Results: *Nosema pyrausta* was deposited in the GenBank under accession number GCA_001924031.1. The approximate genome size was 15.9 Mb, the number of predicted genes was 4468, which is very close to the corresponding indicators of the closely related species *N. bombycis* (Table 1).

Table 1. Main characteristics of the genome of three closely related species of microsporidia of the genus *Nosema*

Metrics	*N. bombycis*	*N. antheraeae*	*N. pyrausta*
GenBank accession number	NA30919	NA183977	GCA_001924031.1
Isolate name	NbCQ1	YY	Npyr2020
Genome size, Mb	15.7	6.6	15.9
N50	57394	1883	8050
Number of genes	4458	3413	4468
GC, %	31	28	34.5
Number of contigs	3558	?	4115
Number of scaffolds	1607	6215	1880
Link	Pan et al., 2013	Pan et al., 2013	Own data

The *Nosema* sp. NspHA22 genome assembly and annotation are not finished by the time of submission of the present work due to higher admixture of host genomic DNA, which requires additional work to filter the nonspecific signals out. Anyway, the raw whole genome data from both microsporidian isolates were used to extract nucleotide sequences of individual genes most often used for multilocus genotyping and multigene phylogenetic reconstructions: heat shock protein HSP70 and RNA polymerase RPB1, as well as hexokinase, additionally proposed for inter- and intraspecific differentiation of microsporidia [9]. Analysis of the corresponding nucleotide sequences of the studied microsporidia showed their division into two sister phylogenetic lineages within the genus *Nosema*. One group included *N. bombycis* and NspHA22, the other included *N. pyrausta* and *N. antheraeae*. The similarity for each of the three genes within the group was about 98–99 %, and between groups – 92–93 %. This makes it possible to fairly clearly distinguish between the studied isolates, but requires sequencing of the corresponding loci for the most closely related forms, since the low frequency of nucleotide substitutions excludes the ability to design primers specific to a particular species (isolate).

Conclusion: The relatively large genome size of *N. pyrausta* compared to other *Nosema* species is believed to be such due to a recent genomic duplication event [10]. Consequently, this event affected both species under consideration, and since their range of susceptible hosts differs significantly [11, 12], the information obtained can serve in further search for genetic determinants of virulence of entomopathogenic microsporidia that cause dangerous diseases of cultivated insects and requiring the development of approaches for prevention and treatment [13], as well as promising from the point of view of plant protection [14]. Genomic data will also facilitate the search for variable regions of the genome for express diagnostics, in particular, for the development of species-specific primers.

Funding: The study is supported the Russian Science Foundation project No. 23-16-00262 and State Assignment project No. FGEU-2022-0019 (state registration No. 122032900144-2).

Список литературы/References

1. Becnel J.J., Andreadis T.G. Microsporidia in Insects. In: Microsporidia: Pathogens of Opportunity. 2014;521-570. doi 10.1002/9781118395264.ch21
2. Han B., Weiss L.M. Microsporidia: obligate intracellular pathogens within the fungal kingdom. *Microbiol Spectr*. 2017;5(2). doi 10.1128/microbiolspec.funk-0018-2016
3. Исси И.В. Развитие микроспоридиологии в России. *Вестник защиты растений*. 2020;103(3):161-176
4. Токарев Ю.С. Значение микроспоридий в борьбе с вредными членистоногими. *Защита и карантин растений*. 2007;(12):14-16
5. Wittner M. Historic Perspective on the Microsporidia: Expanding Horizons. In: The microsporidia and microsporidiosis. Washington: ASM Press, 1999;1-6. doi 10.1128/9781555818227.ch1
6. Bojko J. Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. *Trends Parasitol*. 2022;38(8):642-659
7. Malysh S.M. et al. Susceptibility of Bombyx mori larvae to the microsporidium *Nosema bombycis* from the silkworm and *Nosema sp.* from the cotton bollworm. *Plant Protect News*. 2023;4:210-214. doi 10.31993/2308-6459-2023-106-4-16148
8. Sambrook J., Fritsch E.R., Maniatis T. Molecular Cloning: A Laboratory Manual. NY: Cold Spring Harbor Laboratory Press,1989
9. Tokarev Y.S. et al. Ecological vs physiological host specificity: the case of the microsporidium Nosema pyrausta (Paillot) Weiser, 1961. *Acta Biol Sibirica*. 2022;8:297-316. doi 10.14258/abs.v8.e19
10. Pan G. et al. Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. *BMC Genomics*. 2013;14:186

11. Tokarev Y.S. et al. Hexokinase as a versatile molecular genetic marker for Microsporidia. *Parasitology*. 2019;146(4):472-478. doi 10.1017/S0031182018001737
12. Weiser J. Die Mikrosporidien als Parasiten der Insekten. Beihefte zur Angewandten Entomologie. Hamburg: P.P. Parey, 1961
13. Dolgikh V.V. Construction and heterologous overexpression of two chimeric proteins carrying outer hydrophilic loops of Vairimorpha ceranae and Nosema bombycis ATP/ADP carriers. *J Invertebr Pathol*. 2020;171:107337
14. Malysh J.M., Chertkova E.A., Tokarev Y.S. The microsporidium *Nosema pyrausta* as a potent microbial control agent of the beet webworm *Loxostege sticticalis*. *J Invertebr Pathol*. 2021;186:107675. doi 10.1016/j.jip.2021.107675

doi 10.18699/bgrs2024-5.2-09

Сравнительное изучение структуры и таксономии на примере 5 S и 16 S РНК бактерий

Овчинникова Ю.И.[1], Садовский М.Г.[1, 2, 3]*

[1] *Сибирский федеральный университет, Красноярск, Россия*

[2] *Институт вычислительного моделирования КНЦ СО РАН, Красноярск, Россия*

[3] *СНКЦ ФМБА РФ, Красноярск, Россия*

**msad@icm.krasn.ru*

Ключевые слова: нуклеотидная последовательность; частота; упругая карта; кластер

Мотивация и цель: Выявление статистические свойства нуклеотидных последовательностей, отражающие связь структуры и таксономии исследуемых генов, является ключевой задачей молекулярной биологии и биоинформатики. Одним из классических объектов исследования в данной области являются последовательности 16S рРНК, однако существуют и другие ансамбли последовательностей генов рРНК, представляющие не меньший интерес. В рамках данного исследования мы изучали связь двух структур на примере последовательностей 5S РНК и 16S РНК бактерий.

Методы и алгоритмы: В настоящей работе выявлены подобия и различия структур в упомянутых ансамблях, определяемые триплетным составом генов 16S РНК и генов 5S РНК. Ранее [1] было показано, что триплетный состав генетических систем хорошо коррелирует с таксономическим положением носителей соответствующих генов. Однако зачастую видовой состав баз генетических данных (в нашем случае – базы SILVA) весьма смещён: численность генов различных таксономических групп организмов (бактерий) сильно отличается, что может приводить к искажению результатов. Поэтому необходимо провести индексирование баз данных, т. е. удаление части генов из сверхпредставленных групп. Тот факт, что таксономия бактерий тесно коррелирует с особенностями последовательностей рРНК бактерий, хорошо известен. Стандартный прием выявления такой связи использует технологию сравнения последовательностей генов, основанную на выравнивании (редакционном расстоянии). В нашей работе мы используем иные методы сравнения генов, не опирающиеся на идею выравнивания.

Результаты: перейдём к описанию самой работы и результатов. Сравнение структуры и таксономии индексированных баз генов 5S и 16S РНК бактерий проводилось с помощью метода упругих карт [2]. Для этого последовательности генов преобразовывались в частотные словари триплетов; частотный словарь (триплетов) – это список всех триплетов, встречающихся в последовательности гена, с указанием их частот встречаемости. Тем самым, каждая последовательность гена преобразуется в точку в многомерном метрическом пространстве частот триплетов, что позволяет задать расстояние между точками естественным образом (например, Евклидово расстояние) и применить множество методов анализа структуры распределения таких точек. В нашей работе изучалась кластеризация генов 5S РНК бактерий в сравнении с аналогичной кластеризацией 16S РНК бактерий. Структуры в распределении таких генов могут определяться

как с помощью классификации, так и с помощью кластеризации (метод упругих карт, метод динамических ядер, он же k-means). В нашей работе используется метод упругих карт для кластеризации генов на основе их частотных словарей триплетов. Частотные словари триплетов представляют собой представление генов в виде последовательностей из трех нуклеотидов (например, ACG, GTC и т. д.), и их частоты встречаемости в геноме. Кластеризация по частотным словарям триплетов проводилась с помощью ПО VidaExpert. Результатом настоящей работы является установление связи кластеризации генов 5S и 16S РНК. Так, возможными исходами являются полностью различные паттерны кластеризации, наблюдаемые на этих двух множествах генов, напротив, заметное их подобие. Ещё одним важным является ответ на вопрос: верно ли, что триплетный состав этих двух групп генов позволяет отличить гены 16S РНК и 5S РНК бактерий.

Comparative study of structure and taxonomy interplay over 5 S and 16 S ribosomal genes of bacteria

Ovchinnikova Y.[1], Sadovsky M.[1, 2, 3]*

[1] *Institute of computstional modelling SB RAS, Kasnoyarsk, Russia*

[2] *FSR&CC of FMBA of Russia, Krasnoyarsk, Russia*

[3] *SFU, Krasnoyarsk, Russia*

* *msad@icm.krasn.ru*

Key words: nucleotide sequence; frequency; elastic map; cluster

Motivation and Goal: Revealing the statistical properties of nucleotide sequences that reflect the relationship between the structure and taxonomy of the genes under study is a key task in molecular biology and bioinformatics. One of the classic objects of research in this area is 16S rRNA sequences, but there are other ensembles of rRNA gene sequences that are of no less interest. In this study, we studied the relationship between the two structures using the example of bacterial 5S RNA and 16S RNA sequences. Here we studied the clustering of bacterial 5S RNA genes in comparison to the similar clustering of bacterial 16S RNA. Structures in the distribution of such genes can be determined using both classification and clustering (elastic map method, or k-means). Our work uses the elastic map method to cluster genes based on their triplet frequency triplet ensembles. Triplet frequency dictionaries are a representation of genes as a set of triplets (e.g., ACG, GTC, etc.) provided with their frequency of occurrence in a genome. Clustering by frequency dictionaries of triplets was carried out using VidaExpert software. The result of this work is the connection between the clustering of 5S and 16S RNA genes. Thus, possible clustering outcomes are completely different for these two sets of genes, noticeable similarity in some issues also has been observed. Another important point is whether it is possible to distinguish the 16S RNA and 5S RNA genes of bacteria. The answer is positive on this question.

Список литературы/References

1. Sadovsky M., Putintseva Yu., Chernyshova A., Fedotova V. Genome Structure of Organelles Strongly Relates to Taxonomy of Bearers. In: Ortuño F., Rojas I. (Eds.). Bioinformatics and Biomedical Engineering. IWBBIO 2015. Lecture Notes in Computer Science. Vol. 9043. Springer, 2015;481-490
2. Gorban A., Kégl B., Wunsch D.C., Zinovyev A. (Eds.). Principal Manifolds for Data Visualizations and Dimension Reduction. Springer, 2008

doi 10.18699/bgrs2024-5.2-10

Материалы по истории развития *Populus alba*, произрастающего в европейской части России и на Кавказе, основанные на результатах полногеномного секвенирования

Попченко М.[1]*, Гладыш Н.[1], Володин В.[1, 2], Богданова А.[1], Краснов Г.[1, 2], Кудрявцева А.[1, 2]

[1] *Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

[2] *Центр высокоточного редактирования генома и генетических технологий для биомедицины, Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия*

** popchenko_m@inbox.ru*

Ключевые слова: *Populus alba*; полногеномное секвенирование; филогенетика; популяционная генетика; эволюция

Мотивация и цель: Белый тополь (*Populus alba* L.), эволюционно молодой вид двудомных древесных растений, обычно встречается в умеренных и субтропических регионах Евразии и Северной Африки. В России его можно встретить на юге европейской части, юге Западной Сибири и на Кавказе. В ареале между европейской и кавказской частями существует разрыв. Несмотря на экономическую значимость этого вида и его большой потенциал для использования, молекулярные исследования филогеографии популяций белого тополя в России не проводились. Цель данного исследования – представить гипотезу о формировании природных и культурных ареалов белых тополей в европейской части России.

Методы и алгоритмы: Мы секвенировали полные геномы 36 растений белого тополя, в том числе 12 из Москвы и прилегающих к ней регионов, 2 из Нижегородской, 6 из Волгоградской областей, 9 с Северного Кавказа и 2 из Дагестана. Мы также добавили 3 растения из коллекции Центрального сибирского ботанического сада (местная популяция белых тополей, гибриды между белым тополем и осиной – *Populus canescens* (Aiton) Sm. и var. ‘Хоперский’) и 2 растения, собранные близ Алма-Аты в Казахстане. Секвенирование проводили с использованием платформы Illumina NextSeq 500 (в режиме 2×150 нукл.) с покрытием в диапазоне от 7x до 20x. Данные секвенирования опубликованы в NCBI BioProject (PRJNA1109753). Филогенетический анализ был проведен на основе сравнения профилей замен (SNP), включая точечные SNV и вставки/делеции (indels). Такой анализ позволяет учесть возможность полиплоидии. Прочтения были картированы на референсный геном белого тополя (GCF_005239225) с использованием BWA-MEM (-k 14) с удаленными вторичными выравниваниями [1]. Для поиска замен использовался freeBayes. Далее для векторов частот вариантных аллелей (VAF) было рассчитано попарное евклидово расстояние между образцами. Затем образцы были кластеризованы с использованием алгоритма UPGMA с бутстреппингом. Оценка смешения популяций проводилась с использованием NGSAdmix на основе оценки вероятности генотипирования для всех SNP (-K 3) [2]. Этот метод позволяет оценить наличие генетических примесей в популяциях даже на данных с небольшой глубиной

секвенирования. Для повышения точности в анализ были включены дополнительные образцы деревьев белого тополя, произрастающие в китайской провинции Синьцзян и в Италии (данные получены из NCBI SRA).

Результаты: На основе анализа схожестей профилей SNP-профилей можно выделить по крайней мере три генетически сходные группы: образцы из Нижегородской, Волгоградской областей и Новосибирска образуют кластер (C1), в то время как образцы с Северного Кавказа и Южного Дагестана образуют другой кластер (C2). Образцы из Москвы и прилегающих районов образуют третий кластер (C3). Хотя дендрограммы генетического сходства для хлоропластных, митохондриальных и ядерных геномом различаются, на всех дендрограммах наблюдаются три крупных кластера, причем два кластера в основном состоят из деревьев центрального региона, а еще одна однородная группа включает деревья Сочи, Дагестана и Северо-Кавказского федерального округа (рис. 1).

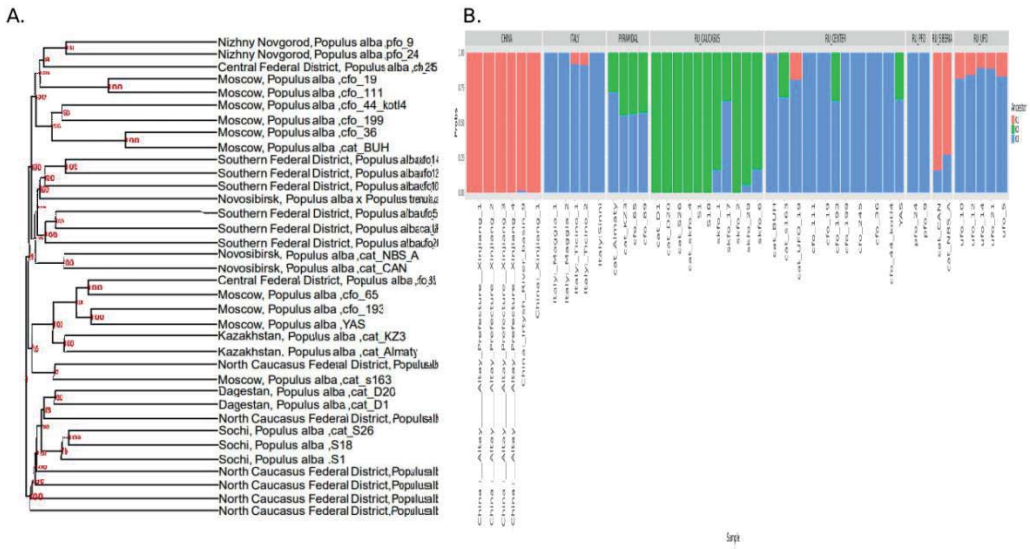


Рис. 1. А – филогенетический анализ деревьев белого тополя, произрастающего на территории России и Казахстана. «Удлиненные ветви» на дендрограмме являются следствием неточностей в определении значений VAF, которые рассчитываются по числу прочтений, подкрепляющих референсную и альтернативную аллель. В – оценка смеси условных популяций для тополей из разных регионов. Некоторые образцы опущены

Проведенный анализ смешения условных популяций белого тополя (NGSAdmix) был основан на гипотезе о существовании трех основных центров, которые могли послужить основой для формирования современной популяции белого тополя на юге европейской части России в послеледниковый период. К таким центрам относятся: Южная Европа, Кавказский регион, а также южные районы Западной Сибири и Южного Урала. В период максимального оледенения на этих территориях сохранялась лесная растительность. Генетический материал из Южной Европы преобладает в геномах популяций Центральной Европы и в лесных районах России. В степях Волгоградской области до четверти генома исследованных образцов содержит генетический материал из Южной Сибири. Следует отметить, что небольшое количество генетического материала из

Западной Сибири присутствует в популяциях Центральной Европы. Также генетический материал из Южной Европы был обнаружен в популяциях Западной Сибири, что вполне естественно, учитывая непрерывное распространение видов в доледниковый период. Кавказский генетический материал полностью формирует геном растений в популяциях Закавказья и Северного Кавказа. Большинство образцов содержат в своих геномах южноевропейский генетический материал. Наличие этого генетического материала можно объяснить историческими связями между Европой и Кавказом через Малую Азию, где белый тополь растет и по сей день, а также миграцией его носителей с севера на юг через долины рек Дон и Кубань. Интересной особенностью является полное отсутствие южносибирского генетического материала у европеоидных популяций, и наоборот. По-видимому, существует разрыв в распространении белого тополя, поскольку он произрастает в восточном Туркменистане, северо-восточном Иране, равнинном Узбекистане и северо-западном Афганистане, что служит препятствием для обмена генетическим материалом между этими регионами. Точное распространение белого тополя в Центральной Азии остается спорным из-за недостатка информации о цветоводстве и долгой истории его выращивания в регионе.

Генетический состав популяций растений в центрально-европейской части России требует отдельного рассмотрения. При анализе образцов растений здесь можно обнаружить следы двух из трех очагов. Генетический материал Южной Европы получен как из южнорусской, так и из центральноевропейской популяций. Роль последних, по-видимому, возрастает вблизи крупных городов и в северо-западных регионах, где это связано с периодом появления парков в конце XIX – начале XX в. Кавказский генетический материал, с другой стороны, связан с интродукцией в середине XX в., когда ботанические сады и селекционные центры начали пополнять свои коллекции материалом со всей страны. Особый интерес представляют пирамидальные формы белого тополя в центральной части Европы. Считается, что эти экземпляры произошли от среднеазиатских пирамидальных типов. Однако, как показал анализ, в их составе присутствует только южноевропейский и кавказский генетический материал. Из-за ограниченного объема доступной информации невозможно дать однозначное объяснение этому открытию. Следовательно, необходим более детальный анализ геномных последовательностей образцов пирамидальных форм белого тополя, произрастающих в Южной Европе и Центральной Азии.

Выводы: Современное распространение белого тополя на юге европейской части России сформировалось в послеледниковый период в результате миграции популяций с Балканского полуострова, Восточной и Западной Сибири и Южного Урала. Кавказские популяции белого тополя не способствовали расселению на этой территории в исторические времена. Современные культивируемые белые тополя в европейской части России включают растения различного происхождения (Центральная Европа, Юг России и Кавказ).

Финансирование: Исследование поддержано Российским научным фондом (грант № 22-14-004).

Materials on the history of the development of *Populus alba*, which grows in the European part of Russia and the Caucasus, are based on whole-genome sequencing

Popchenko M.[1]*, Gladysh N.[1], Volodin V.[1, 2], Bogdanova A.[1], Krasnov G.[1], Kudryavtseva A.[1, 2]

[1] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

[2] *Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

** popchenko_m@inbox.ru*

Key words: *Populus alba*; whole-genome sequencing; phylogenetics; population genetics; evolution

Motivation and Aim: The white poplar (*Populus alba* L.), an evolutionarily young species of dioecious woody plant, is commonly found in temperate and subtropical regions of Eurasia and North Africa. In Russia, it can be found in the southern European part, southern Western Siberia, and the Caucasus. There is a gap in the range between the European and Caucasian parts. Despite the economic importance of this species and its great potential for use, no molecular studies have been conducted on the phylogeography of white poplar populations in Russia. This study seeks to present a hypothesis on the formation of natural and cultivated areas of white poplars in the European part of Russia.

Methods and Algorithms: We sequenced the complete genomes of 36 white poplar plants, including 12 from Moscow and its surrounding regions, 2 from Nizhny Novgorod, 6 from Volgograd regions, 9 from the North Caucasus and 2 from Dagestan. We also added 3 plants from the collection of the Central Siberian Botanical Garden (a local population of white poplars, hybrids between white poplar and aspen – *Populus canescens* (Aiton) Sm. and var. 'Khopersky'), and 2 plants collected near Alma-Ata in Kazakhstan. Sequencing was performed using the Illumina NextSeq 500 platform (2x150 bp), with coverage ranging from 7x to 20x for each sample. The sequencing results are published in NCBI BioProject (PRJNA1109753).

A phylogenetic analysis was conducted based on SNPs profiling, including point SNVs and insertions/deletions (indels). Reads were mapped to the white poplar reference genome (GCF_005239225), using BWA-MEM (-k 14) with removed secondary alignments [1]. FreeBayes was used for variant calling. The Euclidean distance between samples was calculated for variant allele frequency (VAF) vectors, and samples were clustered using UPGMA algorithm with bootstrapping, taking into account the possibility of polyploidy. The assessment of population mixing was performed using NGSAdmix software, based on the assessment of the probability of genotypes for all SNPs (-K 3) [2]. This method allows for the evaluation of admixture at a low sequencing depth. For the accuracy of the analysis, additional samples of white poplar trees were used, growing in the Xinjiang province of China and in Italy. The raw genomic data for these samples were obtained from the NCBI SRA.

Results: Based on the analysis of SNP profiles, at least three genetically similar groups can be identified: samples from the Nizhny Novgorod, Volgograd regions and Novosibirsk form a cluster (C1), while samples from North Caucasus and southern Dagestan form another cluster (C2). Samples from Moscow and surrounding areas form a third cluster (C3). Although the patterns of genetic similarity between chloroplast, mitochondrial and nuclear genomes vary, three large clusters are observed in all dendrograms, with two clusters mainly composed of central region trees and one more homogeneous group containing Sochi, Dagestan and the North Caucasus Federal District trees (Fig. 1).

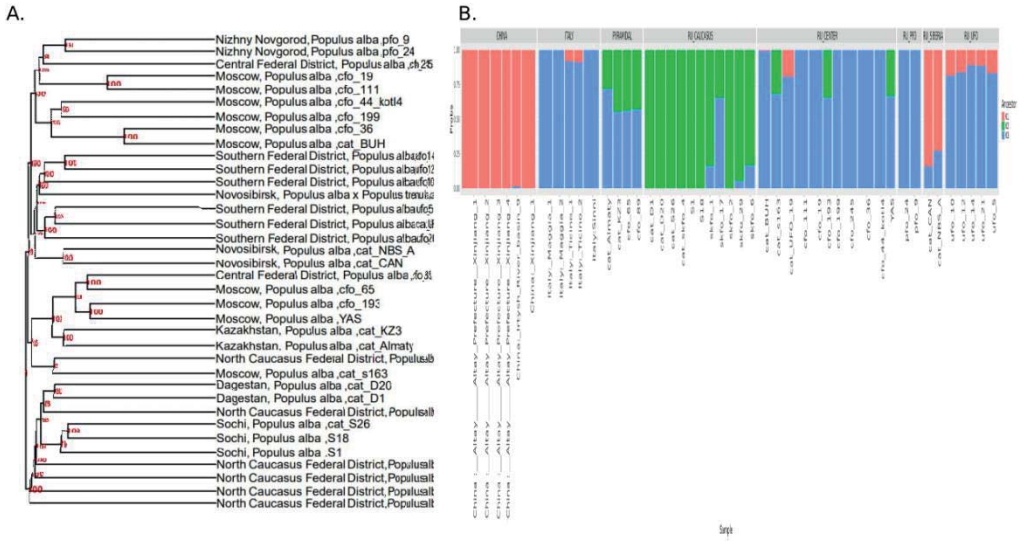


Fig. 1. A, Phylogenetic analysis of white poplar trees growing in Russia and Kazakhstan, with some samples omitted. “Long branches” on the dendrogram are a consequence of inaccuracies in the determination of VAF values, which are calculated by the number of reads supporting the reference and alternative allele. B, Evaluation of a mix of conditional populations for poplars from different regions, with some additional samples also omitted

The population mixing analysis (NGSAdmix) was based on the hypothesis that there were three main centers that could have served as the basis for the formation of the modern white poplar population in the south of the European part of Russia during the post-glacial period. These centers include: Southern Europe, the Caucasus region, and the southern parts of Western Siberia and the Southern Urals. During the period of maximum glaciation, forest vegetation was preserved on these territories. The genetic material from Southern Europe predominates in the genomes of Central European populations and in forest areas of Russia. In steppes of the Volgograd region, up to one-quarter of the genome of the studied samples contains genetic material from South Siberia. It should be noted that a small amount of genetic material from Western Siberia is present in Central European populations as well as genetic material from southern Europe found in populations in Western Siberia, which is quite natural, given the continuous distribution of species during the pre-glacial period. The Caucasian genetic material completely forms the genome of plants in populations of Transcaucasia and the North Caucasus. Most samples contain Southern European genetic material in their genomes. The presence of this genetic material can be explained by historical links between Europe and the Caucasus through Asia Minor, where white poplar still grows today, as well as migration of its carriers from the north to the south via the Don and Kuban River valleys. An interesting feature is the complete absence of South Siberian genetic material from Caucasian populations, and vice versa. There appears to be a gap in the distribution of white poplar since it originated in eastern Turkmenistan, north-eastern Iran, lowland Uzbekistan and north-western Afghanistan, which serves as a barrier to exchange of genetic material between those regions. The exact distribution of white poplar in Central Asia remains debatable due to lack of floral information and the long history of its cultivation in the region.

The genetic composition of plant populations in the central European part of Russia requires separate consideration. When analyzing plant samples, traces of two of the three centers can be found here. Southern European genetic material comes from both Southern Russian and Central European populations. The role of the latter seems to increase near large cities and in Northwestern regions, where it is associated with the period of park introduction in the late 19th–early 20th centuries. Caucasian genetic material, on the other hand, is associated with mid-20th-century introduction, when botanical gardens and breeding centers began expanding their collections with material from all over the country. Pyramidal forms of white poplar in the central part of Europe are of particular interest. It is believed that these specimens were derived from Central Asian pyramidal types. However, as the analysis has shown, only southern European and Caucasian genetic material is present in their composition. Due to the limited amount of information available, it is not possible to provide a definitive explanation for this finding. Therefore, a more detailed analysis of the genome sequences of samples from pyramidal white poplar forms in southern Europe and central Asia would be necessary.
Conclusion: The modern distribution of white poplar in southern European Russia was shaped during the postglacial period through migration from populations in the Balkan Peninsula, eastern and western Siberia, and southern Urals. Caucasus populations of white poplars did not contribute to settlement in this area during historical times. Modern cultivated white poplars in European Russia include plants of diverse origins, such as Central Europe, South Russia, and the Caucasus.
Funding: The study is supported by the Russian Science Foundation (Grant No. 22-14-004).

Список литературы/References

1. Li H., Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760. doi 10.1093/bioinformatics/btp324
2. Skotte L., Korneliussen T.S. et al. Estimating individual admixture proportions from next generation sequencing data. *Genetics.* 2013;195(3):693-702. doi 10.1534/genetics.113.154138

doi 10.18699/bgrs2024-5.2-11

Полиморфные микросателлитные локусы дафний: компьютерное моделирование популяций со смешанной стратегией размножения и их экспериментальный анализ

Порошина А.А.*, Щербаков Д.Ю.
Лимнологический институт СО РАН, Иркутск, Россия
** a.poroshina@lin.irk.ru*

Ключевые слова: Полиморфные локусы; капилярный электрофорез; агентно-ориентированное моделирование; микросателлиты

Мотивация и цель: Полиморфные локусы играют важную роль в изучении генетического разнообразия и эволюционных процессов в популяциях. В частности, для кладоцер [1], таких как дафнии, которые демонстрируют смешанную репродуктивную стратегию, включающую как половое, так и бесполое размножение, понимание полиморфизма локусов может помочь в оценке доли полового размножения и других популяционных параметров. Ранее разработанная агентно-ориентированная модель позволила оценить генетическое разнообразие и параметр неравновесности (D) на смоделированных данных, что открыло возможность использования неравновесности для оценки доли полового размножения в популяции.
Целью настоящей работы является поиск полиморфных локусов в геномах дафний с помощью анализа последовательностей генома и экспериментального анализа реальных проб. Мы используем предсказанные алгоритмом праймеры для амплификации ДНК из проб дафнии и капилярного электрофореза для обнаружения полиморфных локусов [2]. Затем обсуждаются результаты искажения частот микросателлитных аллелей с предсказанными с помощью агентно-ориентированной модели. Таким образом мы оцениваем эффективность метода и уточняем параметры популяции, включая долю полового размножения и эффективный размер популяции.
Методы и алгоритмы: Разработанная ранее агентно-ориентированная [3] модель используется для моделирования перехода популяции диплоидных организмов от полового к бесполому размножению на селективно-нейтральных маркерах. Модель учитывает параметр эффективного размера популяции для расчета θ-параметра с использованием нейтральных маркеров. Специфика микросателлитных маркеров учитывается при моделировании и анализе результатов. Были собраны пробы дафний из различных популяций, была проведена экстракция ДНК из этих проб.
Капиллярный электрофорез был проведен для разделения и анализа амплифицированных ДНК-фрагментов. Результаты капилярного электрофореза были зарегистрированы в виде электрофоретических профилей. Для фрагментного анализа было использовано оригинальное программное обеспечение на языке R. Обнаружены несколько полиморфных маркеров.
Результаты: В рамках настоящей работы мы провели комплексное исследование для поиска полиморфных локусов в геномах дафний.

В результате проведенного исследования, из 33 ранее разработанных микросателлитных маркеров, 10 были идентифицированы как полиморфные в геномах дафний. Эти полиморфные маркеры будут использоваться в дальнейшем для определения стратегии размножения внутри популяций дафний.
Вывод: Настоящая работа демонстрирует успешное применение компьютерного моделирования и экспериментального анализа для поиска полиморфных локусов в геномах дафний. Разработанные полиморфные маркеры будут полезны для оценки доли полового размножения и других популяционных параметров в популяциях дафний. Кроме того, результаты работы открывают перспективы для разработки аналогичных маркеров для других организмов, размножающихся факультативно, таких как гидры. Дальнейшее исследование с помощью этих маркеров может помочь в понимании эволюционных процессов и стратегий размножения в этих организмах.
Результаты настоящей работы могут быть полезны в различных областях, включая экологические исследования, мониторинг и управление популяциями, эволюционные исследования, разработку новых методов и практические приложения.
Финансирование: Исследование было выполнено при финансовой поддержке Российского научного фонда (проект № 22-24-00791 «Метод оценки доли полового размножения у организмов со смешанной репродуктивной стратегией»).

Polymorphic microsatellite loci of daphnia: computer modeling of populations with a mixed breeding strategy and their experimental analysis

Poroshina A.A.*, Sherbakov D.Yu.
Limnological Institute SB RAS, Irkutsk, Russia
** a.poroshina@lin.irk.ru*

Key words: Polymorphic loci; capillary electrophoresis; agent-based modeling; microsatellites

Motivation and purpose: Polymorphic loci play an important role in the study of genetic diversity and evolutionary processes in populations. In particular, for cladocerans [1], such as daphnia, which exhibit a mixed reproductive strategy involving both sexual and asexual reproduction, understanding the polymorphism of loci can help in estimating the proportion of sexual reproduction and other population parameters. The previously developed agent-based model made it possible to estimate genetic diversity and the disequilibrium parameter (D) on the simulated data, which opened up the possibility of using disequilibrium to estimate the proportion of sexual reproduction in a population.
The purpose of this work is to search for polymorphic loci in daphnia genomes using genome sequence analysis and experimental analysis of real samples. We use the primers predicted by the algorithm for DNA amplification from daphnia samples and capillary electrophoresis to detect polymorphic loci [2]. Then, the results of distortion of the frequencies of microsatellite alleles with those predicted using an agent-oriented model are discussed. Thus, we evaluate the effectiveness of the method and refine the population parameters, including the proportion of sexual reproduction and the effective population size.

Methods and algorithms: The agent-oriented model developed earlier [3] is used to simulate the transition of a population of diploid organisms from sexual to asexual reproduction on selectively neutral markers. The model takes into account the parameter of the effective population size to calculate the θ parameter using neutral markers. The specificity of microsatellite markers is taken into account when modeling and analyzing the results. Samples of daphnia from various populations were collected and DNA extraction was performed from these samples.

Capillary electrophoresis was performed to separate and analyze amplified DNA fragments. The results of capillary electrophoresis were recorded in the form of electrophoretic profiles. The original software in the R language was used for the fragment analysis. Several polymorphic markers have been found.

Results: As part of this work, we conducted a comprehensive study to search for polymorphic loci in the genomes of daphnia.

As a result of the study, out of 33 previously developed microsatellite markers, 10 were identified as polymorphic in the daphnia genomes. These polymorphic markers will be used in the future to determine the breeding strategy within daphnia populations.

Conclusion: This work demonstrates the successful application of computer modeling and experimental analysis to search for polymorphic loci in daphnia genomes. The developed polymorphic markers will be useful for estimating the proportion of sexual reproduction and other population parameters in daphnia populations. In addition, the results of the work open up prospects for the development of similar markers for other organisms that reproduce optionally, such as hydra. Further research using these markers may help in understanding the evolutionary processes and reproduction strategies in these organisms.

The results of this work can be useful in various fields, including environmental research, population monitoring and management, evolutionary research, the development of new methods and practical applications.

Funding: The study was carried out with the financial support of the Russian Science Foundation (project No. 22-24-00791 "Method for estimating the proportion of sexual reproduction in organisms with a mixed reproductive strategy").

Список литературы/References

1. Brede N. et al. Microsatellite markers for European Daphnia. *Mol Ecol Notes*. 2006;6(2):536-539
2. Thielsch A. et al. Discrimination of hybrid classes using cross-species amplification of microsatellite loci: methodological challenges and solutions in Daphnia. *Mol Ecol Resour*. 2012;12(4):697-705
3. Poroshina A., Sherbakov D. A Procedure for Modeling Genetic Diversity Distortions in Populations of Organisms with Mixed Reproductive Strategies. *Mathematics*. 2023;11(13):2985

doi 10.18699/bgrs2024-5.2-12

Полногеномное секвенирование энтомопатогенной микроспоридии Tubulinosema loxostegi 2020

Румянцева А.[1]*, Ефейкин Б.[2], Токарев Ю.[1]

[1] *Всероссийский институт защиты растений, Санкт-Петербург, Россия*

[2] *Институт проблем экологии и эволюции им. А.Н. Северцова РАН, Москва, Россия*

** rumiantseva.arina@yandex.ru*

Ключевые слова: микроспоридии; секвенирование; сборка генома

Мотивация и цель: Tubulinosema loxostegi 2020 принадлежит к микроспоридиям, группе облигатных внутриклеточных эукариотических паразитов, родственных грибам. Геномные данные играют важную роль в оценке эволюции вирулентности микроспоридий и поиске взаимосвязи между генетическими различиями разных видов и их гостальной специфичностью [1]. Филогенетический анализ, основанный на гене 16S рРНК, уже недостаточен для идентификации близкородственных видов микроспоридий (сходство данного локуса у представителей рода *Tubulinosema* превышает 99 %) и разрешения более глубоких взаимоотношений таксона для оценки его реального видового разнообразия. В настоящее время доступна сборка геномов примерно 50 видов микроспоридий, однако эта выборка должна быть намного больше, чтобы считаться надежной и полезной, учитывая известное количество адекватно описанных видов, родов и таксонов более высокого ранга [2]. Целью данной работы было секвенирование генома T. *loxostegi* 2020 для сравнения его основных характеристик с единственным представленным на сегодняшний день референсным геномом родственного вида T. ratisbonensis.

Методы и алгоритмы: Микроспоридия *Tubulinosema loxostegi* 2020 была выделена из лугового мотылька, отловленного в Новосибирской области в 2020 г., и размножена в гусеницах сибирского шелкопряда (Dendrolimus sibiricus), содержащихся в лабораторных условиях. Для экстракции геномной ДНК клеточную стенку спор разрушали при помощи жидкого азота. На платформе NovaSeq 4000 Illumina для нашего генома было получено порядка 40 млн прочтений, контроль качества которых осуществлялся при помощи программы FastQC. Выровненные и очищенные от адаптеров прочтения не короче 55 нуклеотидов были использованы для геномной сборки на ассемблере SPAdes. Расчет основных метрик сборки и поиск генов проводили при помощи QUAST и Augustus.

Результаты: Размер полученного генома составил 7 542 987 п.о., распределенных по 1162 контигам с общим содержанием GC 23,12 %, показатель N50 составил ~15 Kb. Размер генома (7.5 Mb), общее количество генов (3086), количество генов, кодирующих белки (3015), рРНК (8) и тРНК (58) практически совпали с таковым для T. *ratisbonensis* (табл. 1), что вполне ожидаемо для двух близкородственных видов.

Сравнительный анализ видов был проведен на основании 6 белок-кодирующих генов, перспективных с точки зрения молекулярных маркеров для дифференциации близкородственных видов, а именно гексокиназы, актина, Hsp70, ДНК-

хеликазы, α- и β-тубулина. Сходство нуклеотидных и аминокислотных последовательностей большинства из них колебалось в пределах 98–99 и 98–100 % соответственно. Для гексокиназы и ДНК-хеликазы эти показатели были несколько ниже, что указывает на более высокий полимофризм данных локусов в пределах рода *Tubulinosema* (табл. 2).

Таблица 1. Сравнение сборок *T. loxostegi* 2020 (SPAdes) и *T. ratisbonensis* (IDBA-UD)

Характеристика	T. loxostegi 2020	T. ratisbonensis
Номер доступа в Генбанке	–	GCA_004000155.1
Число контигов	1162	1011
Самый длинный контиг Mb	1.0	–
Размер генома Mb	7.5	7.6
Количество генов	3086	3082
Количество белков	3015	3013
Количество рРНК	8	8
Количество тРНК	58	55
N50, kb	14.981	11.9
N70, kb	6.887	–
GC, %	23.12	23
Покрытие генома	180.0x	200.0x
Количество неопределенных пар оснований (N's)	11	–
Доля неопределенных пар оснований в 100 Kb сборки	0.67	–
Ссылка	Данная работа	Polonais et al., 2019 [3]

Таблица 2. Сравнение некоторых генов «домашнего хозяйства» T. loxostegi 2020 и *T. ratisbonensis*

Локус	Сходство нуклеотидной / аминокислотной последовательности гена, %
Гексокиназа (1190 п.о.)	96.64 / 98.48
Актин (1128 п.о.)	98.85 / 98.94
Hsp70 (2002 п.о.)	98.50 / 99.70
α-тубулин (1341 п.о.)	98.36 / 99.11
β-тубулин (1320 п.о.)	98.48 / 100
ДНК-хеликаза (1748 п.о.)	95.88 / 93.48

Выводы: Впервые проведены полногеномное секвенирование и предварительная сборка генома для энтомопатогенной микроспоридии из лугового мотылька. О полноте собранного генома свидетельствует сопоставимость основных характеристики сборки T. loxostegi 2020 и близкородственного вида T. ratisbonensis. Сравнительный анализ шести генов «домашнего хозяйства» позволил выявить несколько перспективных молекулярных маркеров для дальнейшей дифференциации близкородственных видов рода Tubulinosema, в частности гены гексокиназы и ДНК-хеликазы, показавшие более высокие различия в нуклеотидных последовательностях, чем ген мсрРНК.

Финансирование: Исследование поддержано грантом РНФ (№ 23-16-00262).

Whole genome sequencing of Tubulinosema loxostegi 2020

Rumiantseva A.[1]*, Efeikin B.[2], Tokarev Y.[1]

[1] *All-Russian Institute of Plant Protection, St. Petersburg, Russia*

[2] *A.N. Severtsov Institute of Ecology and Evolution, RAS, Moscow, Russia*

* *rumiantseva.arina@yandex.ru*

Key words: microsporidia; sequencing; genome assembly

Motivation and Aim: Tubulinosema loxostegi 2020 belongs to the Microsporidia, a group of obligate intracellular eukaryotic parasites related to fungi. Genomic data play an important role in assessing the evolution of microsporidia virulence and searching for the relationship between genetic differences among different species and their host specificity [1]. Phylogenetic analysis based on the 16S rRNA gene is no longer sufficient to identify closely related microsporidia species (the similarity of this locus in representatives of the genus Tubulinosema exceeds 99 %) and to resolve deeper relationships of the taxon to assess its real species diversity. Genome assemblies of approximately 50 microsporidia species are currently available, but this sample would need to be much larger to be considered reliable and useful given the known number of species, genera, and higher-ranking taxa adequately described to date. [2]. The purpose of this work was to assemble of the T. loxostegi 2020 genome to compare its main characteristics with the only reference genome of the related species T. ratisbonensis presented to date.

Methods and Algorithms: The microsporidium Tubulinosema loxostegi 2020 was isolated from the beet webworm larvae (Loxostegi sticticalis) captured in the Novosibirsk region in 2020. Spores were propagated in the Siberian silkworm (Dendrolimus sibiricus) larvae under laboratory conditions. To extract genomic DNA, the cell wall of the spores was destroyed using liquid nitrogen. Approximately 40 million reads were obtained for our genome on the NovaSeq 4000 Illumina platform, and their quality was assessed using the FastQC program. Reads that were aligned and free of adapters, with a minimum length of 55 nucleotides, were utilized for genome assembly using the SPAdes assembler. The main assembly metrics and gene research were calculated using QUAST and Augustus.

Results: The resulting genome size was 7,542,987 bp, distributed over 1,162 contigs with a total GC content of 23.12 %. The N50 indicator was approximately 15 Kb. The genome size (7.5 Mb), total number of genes (3086), number of genes encoding proteins (3015), rRNA (8), and tRNA (58) closely matched those of T. ratisbonensis (Table 1), as expected for two closely related species.

A comparative analysis of species was conducted based on six protein-coding genes, which show promise as molecular markers for differentiating closely related species. These genes include hexokinase, actin, Hsp70, DNA helicase, α-tubulin, and β-tubulin. The similarity of the nucleotide and amino acid sequences for most of these genes ranged from 98 to 99 % and 98 to 100 %, respectively. However, for hexokinase and DNA helicase, these indicators were slightly lower, suggesting a higher level of polymorphism within the genus Tubulinosema at these locus (Table 2).

Table 1. Comparison of *T. loxostegi* 2020 (SPAdes) and *T. ratisbonensis* (IDBA-UD) assemblies

Metrics	T. loxostegi 2020	T. ratisbonensis
GenBank accession number	–	GCA_004000155.1
Number of contigs	1162	1011
Largest contig, Mb	1.0	–
Total assembly length, Mb	7.5	7.6
Number of genes	3086	3082
Number of proteins	3015	3013
Number of rRNA	8	8
Number of tRNA	58	55
N50, kb	14.981	11.9
N70, kb	6.887	–
GC, %	23.12	23
Genome coverage	180.0x	200.0x
N's	11	–
N's per 100 Kbp	0.67	–
Link	–	Polonais et al., 2019 [3]

Table 2. Comparison of some housekeeping genes of *T. loxostegi* 2020 and *T. ratisbonensis*

Locus	Gene nucleotide/amino acid sequence similarity, %
Hexokinase (1190 bp)	96.64 / 98.48
Actin (1128 bp)	98.85 / 98.94
Hsp70 (2002 bp)	98.50 / 99.70
α- tubulin (1341 bp)	98.36 / 99.11
β- tubulin (1320 bp)	98.48 / 100
DNA helicase (1748 bp)	95.88 / 93.48

Conclusion: For the first time, whole-genome sequencing and preliminary genome assembly were carried out for the entomopathogenic microsporidium from the beet webworm larvae. The completeness of the assembled genome is evidenced by the comparability of the main characteristics of the assembly with those of the closely related species T. ratisbonensis. A comparative analysis of six housekeeping genes made it possible to identify several promising molecular markers for further differentiation of closely related species within the genus Tubulinosema, particularly the hexokinase and DNA helicase genes, which exhibited higher differences in nucleotide sequences than the 16s rRNA gene.

Funding: The study is supported by RSF (# 23-16-00262).

Список литературы/References

1. Willis A.R., Reinke A.W. Factors That Determine Microsporidia Infection and Host Specificity. In: Weiss L.M., Reinke A.W. (Eds.). Microsporidia. Experientia Supplementum. Vol. 114. Springer, 2022;91-114
2. Bojko J. et al. Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. *Trends Parasitol*. 2022;38(8):642-659
3. Polonais V. et al. Draft genome sequence of tubulinosema ratisbonensis, a microsporidian species infecting the model organism drosophila melanogaster. *Microbiol Resour Announc*. 2019;8(31):e00077-19. doi 10.1128/mra.00077-19

doi 10.18699/bgrs2024-5.2-13

Сравнительные характеристики первичной и пространственной структуры генов тРНК архей, прокариот и эукариот

Шумилина Т.Г.[1], Орлов Ю.Л.[1, 3], Лебедев Г.С.[1], Анашкина А.А.[1, 2], Ильичёва И.А.[2]

[1] *ФГАОУ ВО Первый Московский государственный медицинский университет имени И.М. Сеченова Министерства здравоохранения Российской Федерации (Сеченовский Университет), Москва, Россия*

[2] *Институт молекулярной биологии имени В.А. Энгельгардта РАН, Москва, Россия*

[3] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

* *tshumilina@mail.ru*

Ключевые слова: промоторы генов тРНК, TATA-бокс, боксы A и B

Мотивация и цель: Структурное сходство, присущее РНК-полимеразам (Pols) бактерий, архей и эукариот, отражает общее происхождение систем транскрипции у всех организмов на Земле [1]. TATA-связывающий белок (TBP) является компонентом общих факторов транскрипции (GTFs) у всех трех ДНК-зависимых РНК-полимераз эукариот, археи используют его ортолог. Ранее, при анализе базальных промоторов генов, транскрибируемых Pol II, нами была обнаружена характерная сингулярная область, отвечающая за непосредственное взаимодействие с белком TBP [2–4]. Целью данной работы было обнаружение аналогичных сингулярных областей, отвечающих за регуляцию транскрипции генов тРНК у архей, прокариот и эукариот.

Методы и алгоритмы: Были проанализированы текстовые и структурные характеристики фрагментов геномов, содержащих гены тРНК и их upstream-областей для наборов генов тРНК представителей архей: *M. barkeri str. Fusaro* (63 гена тРНК), *H. volcanii DS2* (53 гена тРНК), прокариот: *E. coli str. K-12 substr. MG1655* (89 генов тРНК), *B. subtilis subsp. subtilis str. 168* (86 генов тРНК) и низших одноклеточных эукариот *S. pombe 972h-* (171 ген тРНК). Анализировали последовательности участков, расположенных в позициях от –50 до +30 п.н. относительно старта транскрипции, и в позициях от N-60 до N, где N – координата последнего нуклеотида гена тРНК. Получены профили частот встречаемости нуклеотидов, их лого-представления, а также профили структурных и физико-химических свойств.

Результаты: У всех исследованных организмов были обнаружены две сингулярные области, расположенные внутри генов. Их координаты отвечают расположению боксов A и B. Анализ первичной структуры выявил определенные отличия частот встречаемости G:C и A:T пар, характер которых позволяет определить домен, к которому относится конкретный организм. На профилях структурных параметров одноклеточных эукариот *S. pombe 972h-* заметно некоторое снижение жесткости двойной спирали к изменениям угла Roll на участке от –28 до –36 н.п., что может свидетельствовать о более низких энергетических затратах на структурные перестройки ДНК при взаимодействии с

белком TBP именно в этой позиции. Такого эффекта на профилях представителей архей и прокариот мы не обнаружили.
Выводы: Текстовые и структурные параметры двойной спирали ДНК фрагментов нуклеотидных последовательностей генов тРНК и примыкающих к ним областей выше старта транскрипции (до –50 н.п.) из геномов архей (*M. barkeri str. Fusaro* и *H. volcanii DS2)*, прокариот (*E. coli str. K-12 substr. MG1655* и *B. subtilis subsp. subtilis str. 168)* и низших эукариот (*S. pombe 972h-)* свидетельствует о присутствии двух сингулярных областей внутри генов тРНК. Их нуклеотидные последовательности у представителей трех доменов отличаются незначительно. Особенности пространственной структуры двойной спирали ДНК выше старта транскрипции в геномах представителя эукариот *S. pombe 972h-*, свидетельствует о возможности избирательного связывания белка TBP на участке от –28 до –36 н.п., который можно назвать TATA-боксом.
Финансирование: Исследование поддержано грантом Российского научного фонда № 24-24-00563 на тему «Разработка цифровых образовательных программ в биомедицине».

Comparative characteristics of the primary and spatial structure of tRNA genes in archaea, prokaryotes and eukaryotes

Shumilina T.G.[1]*, Orlov Yu.L.[1, 3], Lebedev G.S.[1], Anashkina A.A.[1, 2], Il'icheva I.A.[2]
[1] *Sechenov Moscow State Medical University (Sechenov University), Moscow, Russia*
[2] *Engelhardt Institute of Molecular Biology, Moscow, Russia*
[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
* *tshumilina@mail.ru*

Key words: tRNA gene promoters; TATA box; boxes A and B

Motivation and Aim: The structural similarity shared by RNA polymerases (Pols) of bacteria, archaea, and eukaryotes reflects the common origin of transcription systems in all organisms on Earth [1]. TATA binding protein (TBP) is a component of common transcription factors (GTFs) in all three DNA-dependent RNA polymerases of eukaryotes, and archaea use its ortholog. Previously, when analysing the basal promoters of genes transcribed by Pol II, we discovered a characteristic singular region responsible for direct interaction with the TBP protein [2–4]. The goal of this work was to discover similar singular regions responsible for the regulation of tRNA gene transcription in archaea, prokaryotes, and eukaryotes.
Methods and Algorithms: The textual and structural characteristics of genome fragments containing tRNA genes and their upstream regions were analyzed for tRNA gene sets of archaeal representatives: M. barkeri str. Fusaro (63 tRNA genes), H. volcanii DS2 (53 tRNA genes), prokaryotes: E. coli str. K-12 substr. MG1655 (89 tRNA genes), B. subtilis subsp. subtilis str. 168 (86 tRNA genes) and lower unicellular eukaryotes S. pombe 972h- (171 tRNA genes). The sequences of regions located at positions from –50 bp were analyzed up to +30 bp relative to the start of transcription, and in positions from N-60 to N, where N is the coordinate of the last nucleotide of the tRNA gene.

Frequency profiles of mononucleotides, their logo representations, as well as profiles of structural and physicochemical properties were obtained.

Results: In all organisms studied, two singular regions located inside genes were discovered. Their coordinates correspond to the location of boxes A and B. Analysis of the primary structure revealed certain differences in the frequencies of occurrence of G:C and A:T pairs, the nature of which makes it possible to determine the domain to which a particular organism belongs. Profiles of structural parameters of unicellular eukaryotes S. pombe 972h-, show a slight decrease in the rigidity of the double helix in response to changes of the Roll angle in the region from –28 to –36 bp, which may indicate lower energy costs for structural rearrangements of DNA when interacting with TBP protein in this particular position. We did not find such an effect on the profiles of representatives of archaea and prokaryotes.

Conclusion: Textual and structural parameters of the DNA double helix of fragments of nucleotide sequences of tRNA genes and their upstream regions from the genomes of archaea (M. barkeri str. Fusaro and H. volcanii DS2), prokaryotes (E. coli str. K-12 substr. MG1655 and B. subtilis subsp. subtilis str. 168) and lower eukaryotes (S. pombe 972h-) indicates the presence of two singular regions within tRNA genes. Their nucleotide sequences differ slightly among representatives of the three domains. Features of the spatial structure of the double helix above the start of transcription in eukaryotes (S. pombe 972h-) indicate the possibility of selective binding of the TBP protein in the region from –28 to –36 bp, which can be called the TATA box.

Funding: The research is supported by a grant from the Russian Science Foundation No. 24-24-00563 on the topic “Development of digital educational programs in biomedicine”.

Список литературы/References:

1. Hamada M., Huang Y., Lowe T.M., Maraia R.J. Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. Mol Cell Biol. 2001;21(20):6870-6881
2. Il’icheva I.A., Khodikov M.V., Poptsova M.S. et al. Structural features of DNA that determine RNA polymerase II core promoter. BMC Genomics. 2016;17(1):973
3. Melikhova A.V., Anashkina A.A., Il’icheva I.A. Evolutionary Invariant of the Structure of DNA Double Helix in RNAP II Core Promoters. Int J Mol Sci. 2022;23(18):10873
4. Savina E.A., Shumilina T.G., Tumanyan V.G., Anashkina A.A., Il’icheva I.A. Core promoter regions of antisense and long intergenic non-coding RNAs. Int J Mol Sci. 2023;24(9):8199

doi 10.18699/bgrs2024-5.2-14

Genomic technologies in the study of the phylogeny of plague pathogen and certification of foci of the Caspian region and Central Asia

Balykova A.N.*, Kovrizhnikov A.V., Eroshenko G.A., Krasnov Ya.M., Kutyrev V.V.
Russian Research Anti-Plague Institute "Microbe", Federal Service for Surveillance in the Sphere of Cosumers Rights Protection and Human Welfare, Saratov, Russia
** alinabalnik@gmail.com*

Key words: *Yersinia pestis*; strains of medieval biovar; molecular-genetic analysis; evolution; phylogenetic analysis

Motivation and Aim: The importance of studying the genomic portrait and evolution of the causative agent of natural foci of a particularly dangerous infectious disease - plague, which left an indelible mark in history and still retains its pandemic potential, is beyond doubt. In this work, *Yersinia pestis* strains from plague foci of the Caspian region and adjacent foci of Central Asia, which are located in four countries: Russia, Kazakhstan, Turkmenistan, and Uzbekistan, were studied. The medieval biovar of the main subspecies of the plague pathogen, *Y. pestis subsp. pestis*, circulates in these foci. This biovar is widespread in 33 out of 45 existing plague foci, which is 93.3% of the total area of foci in the CIS countries. Its strains are highly virulent and epidemically significant [1]. The reasons for the high adaptive characteristics of the medieval biovar and its rapid spread in the twentieth century have not yet been determined. We have previously established that at the beginning of the twentieth century two branches of the medieval biovar, 2.MED1 and 2.MED4, which strains were etiologic agents of plague outbreaks with high mortality in the period 1912–1950, were widespread in the territory of the foci of the Northern Caspian region [2]. The aim of this study was to conduct comprehensive MLVA25-/SNP-/CRISPR typing and to estimate the genotype diversity of *Y. pestis* strains in the foci of the Caspian region and Central Asia.
Methods and Algorithms: 350 strains of *Y. pestis* isolated in 1912–2015 from carriers, vectors and humans in 25 natural plague foci of the Caspian Sea, Caucasus and Central Asia were studied. Whole genome sequencing of *Y. pestis* strains was performed using MGI (DNBSEQ-G50RS) platform with MGIEasy FS DNA Library Prep Set and MGIEasy UDB PF Adapter Kit A according to manufacturer's instruction. Fastp v0.23.3 and Trycycler v0.5.4 were applied for data processing. Fragment sequencing was performed using the ABI PRISM 3500XL platform (Applied Biosystems, USA) with 3500 Series Data Collection Software for data processing. For comparative SNP/VNTR/CRISPR analysis developed author's bioinformatics programs (VNTRfinder v0.4, SNPgenotyper v0.1) were used. The Hunter-Gaston diversity index (D) was used to assess the discriminatory power of the typing method.
Results: According to the results of molecular genetic analysis and phylogenetic reconstruction of 81 strains of *Y. pestis* on the basis of 2360 core SNPs, the population structure and genetic profile of the main populations of *Y. pestis* from 25 plague foci of the Caspian region and Central Asia for the period 1912–2015 were determined. WG-SNP analysis of the core genome of 740 strains of the major main and non-main

subspecies of *Y. pestis* was performed to determine stable SNP markers. Typing was performed using the developed programs SNPgenotyper v0.1 and SimGA v.0.4. 59 polymorphic nucleotides (SNPs) were determined, which determine 23 genovariants spread in different epidemic periods of activity of the studied plague foci (Fig 1.).

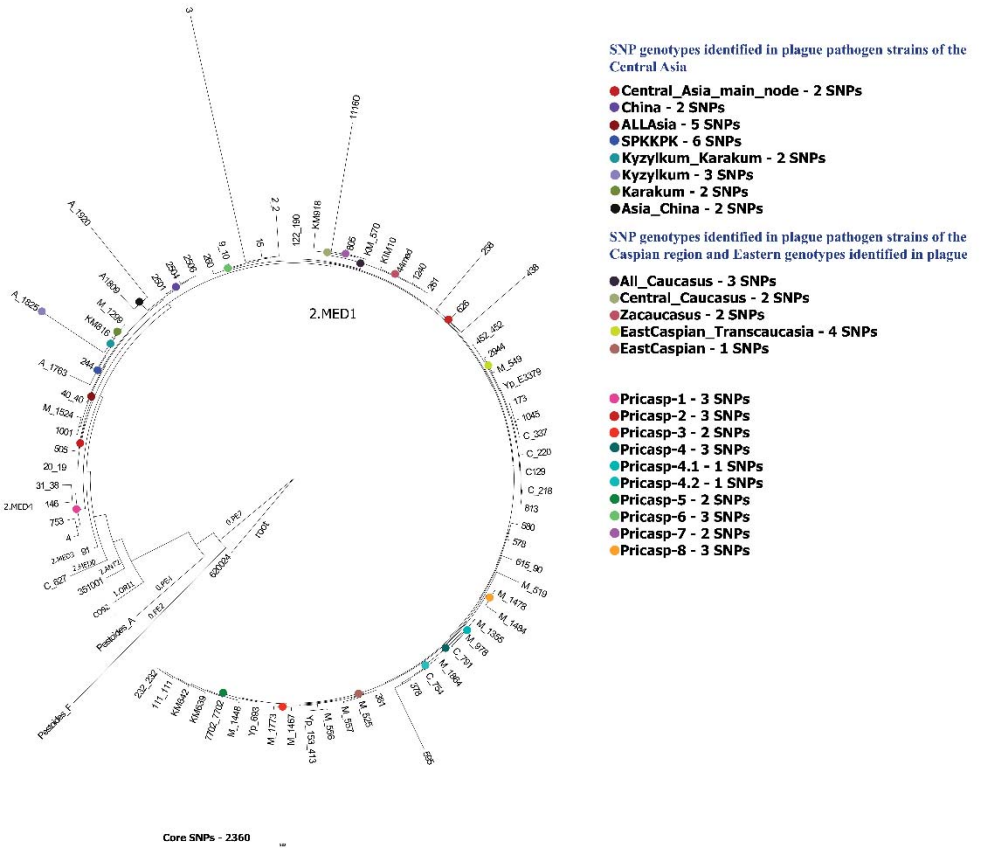


Fig. 1. Population structure of the medieval biovar of *Y. pestis* and established SNP-genotypes of strains from plague foci of the Caspian region and Central Asia of the period 1912–2015.

The Hunter-Gaston discriminating index for this typing scheme was D = 0.9, which ensures the development of SNP panels for the differentiation of populations of the medieval biovar of the plague pathogen. A method of identification of detected SNPs in the plague foci of the Northern and Northwestern Caspian region by fragment sequencing was also developed.

VNTR loci were searched using the MLVA25 scheme and CRISPR loci using the VNTRFinder v0.4 program. Based on the data received VNTR loci (Variable number tandem repeats) with high resolution for differentiating these populations were identified. The loci ms05, ms06, ms07, ms15, ms46, ms56, ms62, ms70, ms74 (number of locus alleles > 3, allele polymorphism index h > 0.5) had the highest variability. Comparative analysis of the locus composition of CRISPR elements, which are

responsible for the immune system of bacteria and can store information about the past interventions of phages, was carried out for the loci characteristic of the plague pathogen, YPa (YP1), YPb (YP2), and YPc (YP3) [3]. The vast majority had YPa (YP1) size of 305 bp, YPb (YP2) locus had a size of 321 bp, and YPc (YP3) locus had a size of 321 bp. The CRISPR loci possessed the following spacer profile: YPa a1-a2-a3; YPb b1-b2-b3-b3-b4`; YPc c1-c2-c3. This profile was identified previously as characteristic of the medieval biovar of the main subspecies [4]. Some strains had a unique spacer profile, which can be further used as a marker in genetic certification of foci of the plague pathogen.

Conclusion: Based on the results of a comprehensive analysis of 350 strains of *Y. pestis* isolated in 1912–2015 in 25 natural plague foci in four countries: Russia, Kazakhstan, Turkmenistan and Uzbekistan, new data on the evolution and genetic variability of populations of the highly virulent and epidemically significant medieval biovar of the plague pathogen were obtained. SNP-/VNTR-/CRISPR loci of the genome of the medieval biovar were analyzed using the author's computer programs. 59 SNPs were identified and SNP panels are being developed for molecular genetic certification of plague foci. The profile of CRISPR elements and spacers (YPa a1-a2-a3; YPb b1-b2-b3-b4`; YPc c1-c2-c3) was established; VNTR loci providing differentiation of genetic populations of *Y. pestis* medieval biovar of different origin were determined. The obtained results are important for the study of variability of the genome architecture of the plague pathogen and for improving the efficiency of molecular epidemiologic tracing of the origin of pathogenic strains during molecular epidemiologic monitoring of plague foci.

Funding: This work did not receive any financial support from commercial or non-commercial organizations.

References

1. Kutyrev V.V. et al. Phylogeny and Classification of *Yersinia pestis* Through the Lens of Strains From the Plague Foci of Commonwealth of Independent States. *Front. Microbiol*. 2018;9:1106. doi 10.3389/fmicb.2018.01106
2. Eroshenko G.A. et al. Evolution and circulation of *Yersinia pestis* in the Northern Caspian and *Northern* Aral Sea regions in the 20th–21st centuries. *PLoS One*. 2021 Feb 11;16(2):e0244615. doi 10.1371/journal.pone.0244615
3. Pourcel C., Salvignol G., Vergnaud G. Pourcel C. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*. 2005;151(Pt 3):653-663. doi 10.1099/mic.0.27437-0
4. Cui Y. et al. Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced palindromic repeats. *PLoS One*. 2008;3(7):e2652. doi 10.1371/journal.pone.0002652

doi 10.18699/bgrs2024-5.2-15

Culture independent barcoding of naked lobose amoebae (Amoebozoa: Tubulinea, Discosea, Variosea) from freshwater and soil habitats

Glotova A.[1, 2]*, Nassonova E.[1, 2], Abakumov E.[3], Smirnov A.[1, 2]

[1] *Laboratory of Cytology of Unicellular Organisms, Institute of Cytology, RAS, St. Petersburg, Russia*

[2] *Department of Invertebrate Zoology, Faculty of Biology, Saint Petersburg State University, St. Petersburg, Russia*

[3] *Department of Applied Ecology, Faculty of Biology, Saint Petersburg State University, St. Petersburg, Russia*

* *glotova.anna@gmail.com*

Key words: Amoebozoa; Tubulinea; Discosea; Variosea; biogeography; metabarcoding

Motivation and Aim: Within the complex framework of the protist distribution patterns debate [1] naked lobose amoebae (NLA) are considered one of the most challenging phylogenetic group. Despite their abundance in many types of natural habitats [2], they are rarely accurately described in ecological studies due to the lack of easily assessed identification characters [3]. Since the conventional two-stepped molecular barcoding approach [4, 5] often requires a labor-consuming stage of amoebae cultivation [6], it has been still applied only for some culturable species and, in the recent decade – to single cells isolated from environment [7]. NGS metabarcoding provides a new tool to describe undiscovered amoebae diversity and could be applied for numerous minimally processed environmental samples.

Methods and Algorithms: Total environmental DNA was extracted from 9 freshwater bottom sediment samples (Turovskoe lake and stream between Turovskoe and Zaklynskoe lakes, Northwestern Russia) and 2 soil samples (Tomsk, Western Siberia, Russia and Etna, Italy) using Dneasy PowerLyzer PowerSoil DNA isolation Kit (Qiagen). Primers biased against metazoans EUK565F_NGS/UnonMet_R [8] and Q5 High Fidelity DNA Polymerase Kit (New England Biolabs) were used in order to amplify V4 marker region of 18S rRNA gene.

Amplicon library construction and sequencing using Illumina MiSeq platform were provided by the Core Facility Center “BioBank” of the Research Park of St. Petersburg State University (https://researchpark.spbu.ru/en/biobank-eng). After quality control reads were assembled into contigs using Mothur (https://mothur.org/wiki/MiSeq_SOP) [10], the same software was also used for data proceeding up to unique sequences insertion into the eukaryotic reference alignment from the SILVA database (https://mothur.org/wiki/silva_reference_files/) and their classification. Filtered target sequences and unclassified sequences were aligned using MAFFT v. 7.490 (https://mafft.cbrc.jp/alignment/server) [11], taxonomically attributed using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and added into local alignments of main phylogenetic lineages of NLA. SeaView v. 4.6.1 [12] was used for alignments editing. ML analysis was performed by IQ-TREE v1.6.12 [13] with the model GTR+I+G+F, resulting phylogenetic trees were visualized using iTOL v.3 (http://itol.embl.de) [14].

Results: High-level taxonomic profiling showed a strong prevalence of Opisthokonta (Ascomycota, Basidiomycota, Mucoromycota, Chytridiomycota) in soil samples and SAR (Ciliophora, Chrysophyta, Bacillariophyta) in freshwater bottom sediment samples. Though automatic OTUs classification allowed to recover particular Variosea (Schizoplasmodiidae) from different samples, most of Amoebozoa sequences were obtained from unclassified reads array.

Prevalent amount of NLA sequences, obtained from almost every analyzed soil and freshwater sample, corresponded to Variosea (291 reads affiliated to *Filamoeba*, *Arboramoeba*, *Heliamoeba*, *Flamella*, *Telaepolella*, *Ischnamoeba*, *Dictyamoeba*, *Angulamoeba*) and Tubulinea (435 reads affiliated to *Flabellula*, *Leptomyxa*, *Copromyxa*, *Ptolemeba*, *Hartmannella*, *Echinamoeba*, *Vermamoeba*).

Less numerous target sequences appeared within Cutosea (70 reads affiliated to *Squamamoeba*, *Phreatamoeba*, *Mastigamoeba*) and Flabellinia (61 reads affiliated to *Mycamoeba*, *Vannella*, *Korotnevella*). Representatives of Longamoebia (15 reads affiliated to *Dermamoeba*) were obtained only from two freshwater samples (T.06.20, P.08.21).

Recovered diversity of several widespread genera commonly used for culture-based barcoding in laboratory conditions (*Vannella*, *Korotnevella*, *Mayorella*) appeared surprisingly limited. At the same time metabarcoding data visibly contributed into the diversity of *Filamoeba*, *Angulamoeba*, *Echinamoeba*, *Vermamoeba*, *Mycamoeba*, though these amoebae are rarely cultivated and assessed in traditional ecological studies.

A noticeable proportion of newly obtained sequences clustered only with environmental Amoebozoa sequences deposited in Genbank with no morphological data affiliated. Patchy and sporadic patterns of sequences distribution within analyzed freshwater and soil samples might reflect population explosion events in certain conditions on a microhabitat level.

Conclusion: Metabarcoding data on NLA from freshwater bottom sediments and soil allowed to identify representatives of all major phylogenetic groups. Among them there are numerous unculturable species, unavailable for culture-based barcoding. Since a significant part of natural diversity of NLA still remains underdescribed and their communities are likely to be sensitive to local conditions favorability, metabarcoding appears particularly promising and complements traditional culture-based methods in amoebae ecological studies.

Funding: The study was supported by the Russian Science Foundation (project No. 23-74-00050) and used equipment of the core facility centers “Culturing of microorganisms”, “Development of molecular and cell technologies” and “Biobank” of the Research Park of Saint Petersburg State University.

References

1. Foissner W. Protist diversity and distribution: Some basic considerations. *Biodivers Conserv.* 2008;17(2):235-242. doi 10.1007/s10531-007-9248-5
2. Smirnov A., Thar R. Spatial distribution of gymnamoebae (Rhizopoda, Lobosea) in brackish water sediments at the scale of centimeters and millimeters. *Protist*. 2003;154:359-369. doi 10.1078/143446103322454121
3. Smirnov A., Chao E., Nassonova E., Cavalier-Smith T. A revised classification of naked lobose amoebae (Amoebozoa: Lobosa). *Protist.* 2011;162:545-570. doi 10.1016/j.protis.2011.04.004
4. Pawlowski J., Audic S., Adl S., Bass D. et al. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* 2012;10(11):e1001419. doi 10.1371/journal.pbio.1001419

5. Kudryavtsev A. Amoebozoan barcoding marker cytochrome c oxidase (Cox1), RNA editing and issues in creating a public reference sequence database. *Protistology*. 2022;16(3):236-245. doi 10.21685/1680-0826-2022-16-3-8
6. Zlatogursky V., Kudryavtsev A., Udalov I., Bondarenko N., Pawlowski J., Smirnov A. Genetic structure of a morphological species within the amoeba genus *Korotnevella* (Amoebozoa: Discosea), revealed by the analysis of two genes. *Eur J Protistol.* 2016;56:102-111. doi 10.1016/j.ejop.2016.08.001
7. Mesentsev Y., Bondarenko N., Kamyshatskaya O. et al. Thecochaos is not a myth: study of the genus Thecochaos (Amoebozoa, Discosea) – a rediscovered group of lobose amoeba, with short SSU gene. *Org Divers Evol* 2023;23:7-24. doi 10.1007/s13127-022-00581-9
8. Bass D., Del Campo J. Microeukaryotes in animal and plant microbiomes: Ecologies of disease? *Eur J Protistol.* 2020;76:125719. doi 10.1016/j.ejop.2020.125719
9. Weiss L., Vossbrinck C. Molecular biology, molecular phylogeny and molecular diagnostic approaches to the Microsporidia. In: Wittner M., Weiss L.M. The Microsporidia and Microsporidiosis. Amer Society for Microbiology, 1999. doi 10.1128/9781555818227.CH4
10. Schloss P. Reintroducing mothur: 10 years later. *Appl Environ Microbiol.* 2020;86:e02343-19. doi 10.1128/AEM.02343-19
11. Katoh K., Rozewicki J., Yamada K. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinf.* 2019;20(4):1160-1166. doi 10.1093/bib/bbx108
12. Gouy M., Guindon S., Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 2010;27(2):221-224. doi 10.1093/molbev/msp259
13. Nguyen L., Schmidt H., von Haeseler A., Minh B. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268-274. doi 10.1093/molbev/msu300
14. Letunic I., Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016;44(W1):W242-W245. doi 10.1093/nar/gkw290

doi 10.18699/bgrs2024-5.2-16

Pangenome analysis of *Stenotrophomonas maltophilia* antiphage defence systems

Jdeed G.[1]*, Morozova V.V.[2], Tikunova N.V.[2]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

** ghadeerjdeed@outlook.com*

Key words: antiphage; defence systems; *Stenotrophomonas maltophilia*; pangenomics

Motivation and Aim: *S. maltophilia* is a highly heterogeneous species known for its ability to adapt to various environments [1, 2]. The study aimed to investigate the distribution and conservation of antiphage defense systems in five newly isolated *S. maltophilia* strains from different ecosystems in Novosibirsk, Russia, and compare them with other available *S. maltophilia* strains in the NCBI GenBank.

Methods and Algorithms: The genomes of the isolated strains were assembled and processed using SPAdes [3] and Ragtag [4] among other tools, searching for antiphage defense systems was done using defense-finder [5], determining conservation was done using MUSCLE alignment [6] with additional calculations, searching for prophages and CRISPR/Cas systems was done using Phigaro [7] and CRISPR/Cas Finder [8], respectively, pangenome analysis was done mainly using Roary [9].

Results: The results showed that the five locally sequenced *S. maltophilia* strains shared 65 % of their genes, compared to 4 % shared among most strains in the pangenome matrix. Only 8 % of the studied strains had complete CRISPR/Cas systems, which is lower than the general distribution in bacteria (42 %) [10]. A total of 72 different antiphage systems and subsystems were identified, with the anti ssDNA phages Wadjet I system, restriction modification systems, Gabija system with endonuclease activity, and the abortive infection system AbiE 4 being the most common (Fig. 1). The distribution of these systems did not follow a trend related to the site of isolation of the strains or their evolutionary history, suggesting that the acquisition of these systems is linked to the presence of phages and other bacterial strains in the microecosystems. Most of the widely distributed systems were highly variable, except for Wadjet I and AbiE, which were conserved, the essential *JetD* gene of the topoisomerase IV activity was absent from half Wadjet I systems, suggesting that those systems are inactive [17] (Fig. 2).

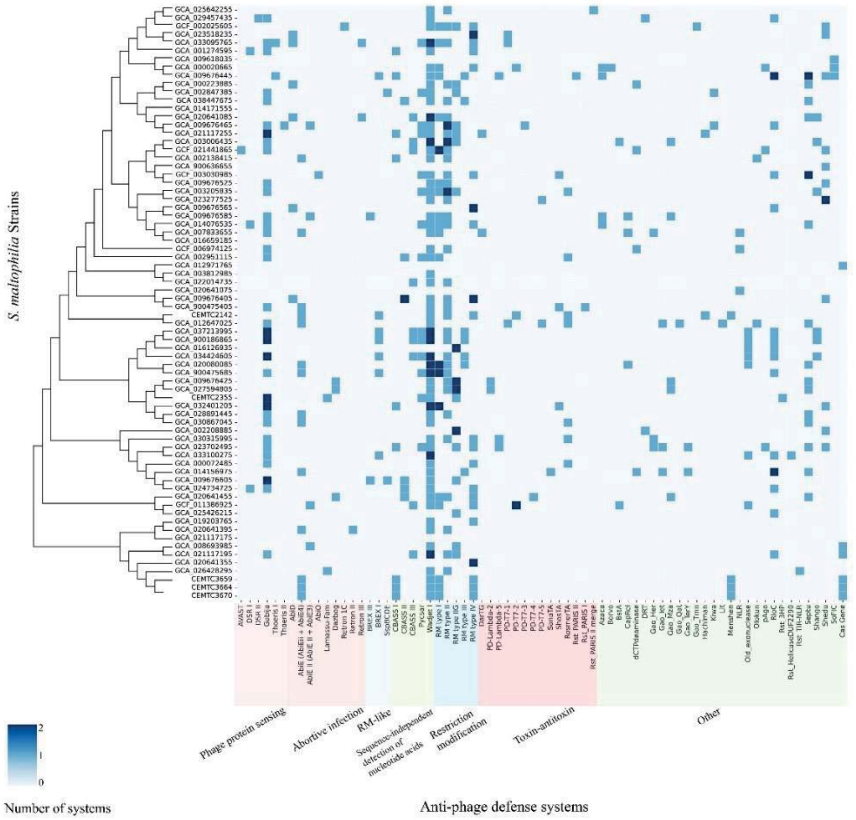


Fig. 1. Antiphage defense systems found in 71 *S. maltophilia* strains and their classification

Fig. 2. Conservation of *S. maltophilia* antiphage defense systems

Conclusion: Antiphage defense systems are widely variable among *S. maltophilia* strains, studying them helps determining one of the factors that determine the efficiency of utilizing phages in therapy.

Funding: This study was supported by the Ministry of Science and Higher Education of the Russian Federation under project grant No. 075-15-2021-1085.

References

1. Palleroni N.J., Bradbury J.F. Stenotrophomonas, a new bacterial genus for Xanthomonas maltophilia (Hugh 1980) Swings et al. 1983. *Int J Syst Bacteriol*. 1993;43:606-609. doi 10.1099/00207713-43-3-606
2. Van den Mooter M., Swings J. Numerical analysis of 295 phenotypic features of 266 Xanthomonas strains and related strains and an improved taxonomy of the genus. *Int J Syst Bacteriol.* 1990;40:348-369. doi 10.1099/00207713-40-4-348
3. Bankevich A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19:455-477. doi 10.1089/cmb.2012.0021
4. Alonge M., Lebeigle L. et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol*. 2022;23:258. doi 10.1186/s13059-022-02823-7
5. Tesson F., Hervé A. et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat Commun.* 2022;13:2561. doi 10.1038/s41467-022-30269-9
6. Madeira F., Pearce M. et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 2022;50:W276-W279. doi 10.1093/nar/gkac240
7. Starikova E.V., Tikhonova P.O. et al. Phigaro: High-throughput prophage sequence annotation. *Bioinformatics.* 2020;36:3882-3884. doi 10.1093/bioinformatics/btaa250
8. Couvin D. et al. Pourcel C. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for cas proteins. *Nucleic Acids Res.* 2018;46:W246-W251. doi 10.1093/nar/gky425
9. Page A.J. et al. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 2015;31:3691-3693. doi 10.1093/bioinformatics/btv421
10. Makarova K.S. et al. Evolutionary classification of CRISPR-cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol.* 2020;18:67-83. doi 10.1038/s41579-019-0299-x

doi 10.18699/bgrs2024-5.2-17

A novel environmental pseudomonas jumbo phage PseuM_296: genome analysis and putative taxonomy

Morozova V.V.[1]*, Mogileva A.A.[1, 2], Kozlova Yu.N.[1], Tikunov A.Yu.[1], Fedorets V.A.[1, 2], Yakubovskij V.I.[1, 2], Zhirakovskaya E.V.[1], Tikunova N.V.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *Faculty of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

* *morozova@niboch.nsc.ru*

Key words: jumbo phage; Pseudomonas mandelii; phiKZ; Caudoviricetes; myovirus

Motivation and Aim: The genus Pseudomonas is a very large taxonomic group, it contains more than three hundred species. They represent a significant part of microbial communities, are both pathogens of plants and animals and their commensals. These bacteria encode a huge variety of metabolic pathways, can produce a number of organic compounds, and are able to disrupt various xenobiotics, therefore, the members of the genus are extensively studied. In contrary, environmental Pseudomonas bacteriophages have been poorly investigated, the vast majority of known Pseudomonas phages are specific to the nosocomial pathogen Pseudomonas aeruginosa. As the environmental Pseudomonas phages can affect pseudomonads, which are used in biotechnology or agriculture, so these phages need to be studied. The phage PseuM_296 and its bacterial host were isolated from the water sample from Kazanka river in Tatarstan Republic, RF. Based on 16S rRNA analysis, bacterial host strain was identified as Pseudomonas mandelii (P. fluorescens lineage) and deposited in Collection of Extremophilic Microorganisms and Type Cultures of ICBFM SB RAS as strain P. mandelii CEMTC 6978. Phage PseuM_296 genome was sequenced, analyzed and its preliminary taxonomic position was identified.

Methods and Algorithms: A negative staining electron microscopy technique was applied to examine phage PseuM_296 particles morphology. The phage suspension (109 pfu/ml) was adsorbed on a copper grid and contrasted using 1 % uranyl acetate; afterwards, the grid was examined for phage particles with a JEM 1400 transmission electron microscope (JEOL, Tokyo, Japan). To determine the sequence of the PseuM_296 genome, DNA was extracted from the phage preparation as described previously [1]. Briefly, phage particles suspension was incubated with RNase and DNase (Thermo Fisher Scientific, USA) for 1 h at 37 ºC. Then, the phage suspension was supplemented with EDTA, proteinase K (Thermo Fisher Scientific, USA) and SDS to final concentrations of 20 mM, 100–200 mkg/ml, and 0.5 %, respectively, and the mixture was incubated for 3 h at 55 °C. After that, phage DNA was purified by phenol/chloroform extraction and subsequent ethanol precipitation. Paired-end library of the PseuM_296 genome was done using the Nextera DNA Sample Preparation Kit (Illumina, Inc, USA). Sequencing was carried out using the MiSeq Benchtop Sequencer and MiSeq Reagent Kit v.1 (2 × 250 base reads) (Illumina Inc, USA). Genome sequence was de novo assembled using the SPAdes genome assembler v.3.15.2 (http://cab.spbu.ru/software/spades). Rapid Annotation Subsystem Technology (RAST) v.2.0 (https://rast.nmpdr.org) was used to predict putative open reading frames (ORFs). Next, obtained ORFs were checked manually against the NCBI GenBank protein

database (https://www.ncbi.nlm.nih.gov) using BLASTX and DELTA-BLAST algorithms. In addition, InterProScan and HHPred software [2, 3] were used to analyze ORFs encoding proteins without homology with the sequences deposited in the GenBank database. A comparative proteomic phylogenetic analysis was performed using Viral Proteomic tree server (ViPTree server) (https://www.genome.jp/viptree). Sequence identity matrix was calculated using Virus Intergenomic Distance Calculator, VIRIDIC (http://rhea.icbm.uni-oldenburg.de/VIRIDIC).

Results: Electron microscopy revealed that the PseuM_296 particle consists of a large capsid (Ø130 nm) connected to a long contractile tail (L = 230 nm). Therefore, the virion morphology corresponds to the myovirus morphotype. The PseuM_296 genomic characteristics were as follows: the genome length was 273 174 bp; it contained 332 putative genes and three of them correspond to tRNAs. Only eighty-one genes encode proteins with predicted functions that were determined based on their amino acid sequences and domain structure similarity. The remaining 248 genes were defined as hypothetical. Putative multisubunit RNAPs (both virion RNA polymerase and nonvirion RNA polymerase), SbcCD complex ATPase, DnaB-like replicative helicase, RNA helicase, and family B DNA polymerase were revealed in the PseuM_296 genome. These genes have been defined previously as the core ones specific to the phiKZ-like group of jumbo phages [4, 5]. In addition, genes encoding chimallin (a major component of the nucleus-like compartment) and PhuZ (tubulin-like protein) were detected, which provide a unique lifestyle of jumbo phages. Phylogenetic proteomic analysis confirmed that phage PseuM_296 belongs to phiKZ-like phages and the most similar phage was Pseudomonas phage pPa_SNUABM_DT01 (Fig. 1).

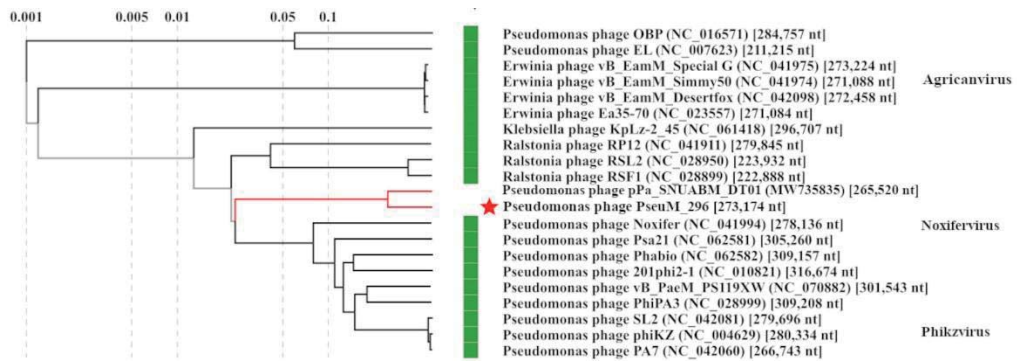


Fig. 1. Phylogenetic proteomic analysis of the jumbo phage PseuM_296. PseuM_296 genome was marked with asterisk. Group of pPa_SNUABM_DT01 and PseuM_296 phages is marked with red

The analysis of intergenomic similarity for PseuM_296 and fourteen most similar phages was carried out using VIRIDIC tool. It has been shown that pPa_SNUABM_DT01 and PseuM_296 genomes have a limited level of similarity (52.4 %), this level is significantly lower than the threshold value for grouping phages into one genus (70 %). Therefore, these phages may be prototype species for two new genera among phiKZ-like phages (Fig. 2).

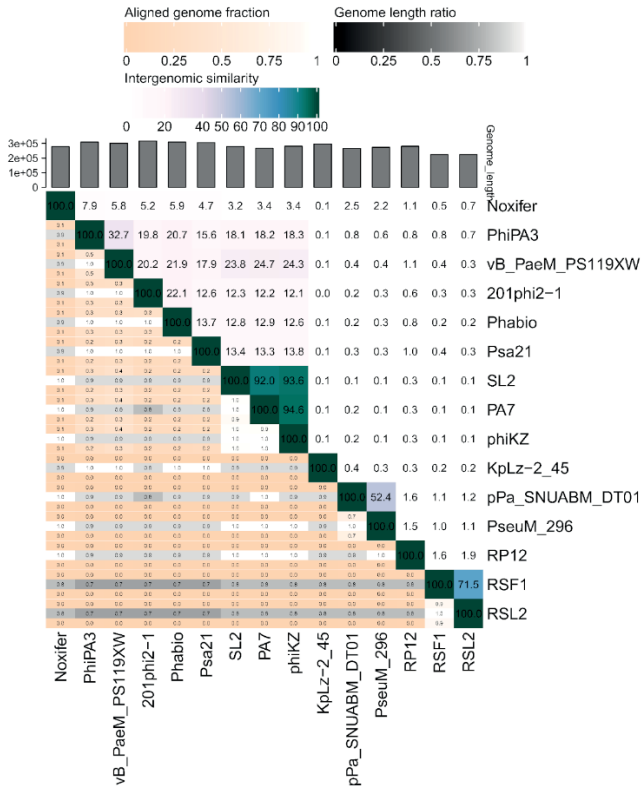


Fig. 2. An intergenomic similarity matrix was calculated for the studied phage PseuM_296 and the phiKZ-like phages closest to it

Conclusion: A novel jumbo phage, PseuM_296, has been isolated, specific to the *P.mandelii* inhabiting in the environment. The particles of phage PseuM_296 had the morphotype of myovirus. Analysis of the PseuM_296 genome has shown that PseuM_296 is a new representative of the phiKZ-like group of giant phages and may be a prototype species of a new genus.

Funding: Funding: This study was supported by the Ministry of Science and Higher Education of the Russian Federation under project grant No. 075-15-2021-1085.

References

1. O'Flaherty S., Coffey A., Edwards R. et al. Genome of staphylococcal phage K: a new lineage of Myoviridae infecting gram-positive bacteria with a low GC content. *J Bacteriol.* 2004;186:2862-2871
2. Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. InterProScan: protein domains identifier. *Nucleic Acids Res.* 2005;33(Web Server issue):W116-W120. doi 10.1093/nar/gki442
3. Söding J., Biegert A., Lupas A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005;33(Web Server issue):W244-W248. doi 10.1093/nar/gki408
4. Sokolova M.L., Misovetc I.V., Severinov K. Multisubunit RNA polymerases of Jumbo bacteriophages. *Viruses.* 2020;12(10):1064. doi 10.3390/v12101064
5. Lavysh D., Sokolova M., Minakhin L., Yakunina M., Artamonova T., Kozyavkin S., Makarova K.S., Koonin E.V., Severinov K. The genome of AR9, a giant transducing *Bacillus* phage encoding two multisubunit RNA polymerases. *Virology.* 2016;495:185-196. doi 10.1016/j.virol.2016.04.030

doi 10.18699/bgrs2024-5.2-18

Metabarcoding study of the phylogenetic diversity of basal Holomycota (Opisthokonta) in the bottom sediments of Turovskoe Lake (Northwestern Russia)

Nassonova E.*, Glotova A.

Laboratory of Cytology of Unicellular Organisms, Institute of Cytology, RAS, St. Petersburg, Russia
** nosema@mail.ru*

Key words: Microsporidia; Rozellida; Aphelida; Holomycota; molecular phylogeny; metabarcoding

Motivation and Aim: Basal Holomycota include three groups of obligate intracellular parasites – Rozellida, Microsporidia and Aphelida [1]. Their representatives are suspected to influence on a function of food webs in different types of aquatic ecosystems [2] and population dynamics of various microeukaryotes [3]. Since NGS provides a new opportunity for extensive study of phylogenetic diversity of unculturable species we obtained metabarcoding data on 9 samples of freshwater bottom sediments from two closely located sampling sites (Turovskoye Lake and the stream between Turovskoye and Zaklinskoe lakes, Leningrad region, Russia) and analyzed them with a special focus on basal Holomycota [4, 5].
Methods and Algorithms: Amplification of the V4 region of 18S rRNA gene of rozellids and aphelids from the total environmental DNA extractions was performed through (1) nested PCR with eukaryotic primers 18sEUK581F/18sEUK1134R [6] and Holomycota-specific primers E572/1009R [7] subsequently or (2) single-step PCR with the primers biased against metazoans EUK565F_NGS/UnonMet_R [8]. Amplification of V1-V3 regions of 18S rRNA gene of Microsporidia was performed through single-step PCR with specific primers 18F/530R or V1F/530R [9]. Q5 High Fidelity DNA Polymerase Kit (New England Biolabs) was used for each amplification. Amplicon library construction and sequencing using Illumina MiSeq platform were provided by the Core Facility Center "BioBank" of the St. Petersburg State University Research Park (https://researchpark.spbu.ru/en/biobank-eng).
Quality-controlled reads were assembled into contigs using Mothur (https://mothur.org/wiki/MiSeq_SOP) [10], with the same software used for data proceeding up to unique sequences insertion into the eukaryotic reference alignment from the SILVA database (https://mothur.org/wiki/silva_reference_files/) and their classification. Filtered target sequences were aligned using MAFFT v. 7.490 (https://mafft.cbrc.jp/alignment/server) [11], taxonomically attributed using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and added into alignments of basal Holomycota (2420 sequences from Genbank). SeaView v. 4.6.1 [12] was used for alignments editing. ML analysis was performed by IQ-TREE v1.6.12 [13] with the model GTR+I+G+F, resulting phylogenetic trees were visualized using iTOL v.3 (http://itol.embl.de) [14].
Results: Each amplification approach resulted in sequencing a significant amount of SAR representatives (Ciliophora, Chrysophyta, Bacillariophyta), some Archaeplastida (Chlorophyta) and non-targeted Opisthokonta (Ascomycota, Basidiomycota, Mucoromycota, Chytridiomycota) from all freshwater samples. Both PCR protocols

used to obtain rozellids and aphelids (nested PCR with 18sEUK581f/18sEUK134r and E572/1009R primers subsequently and one-step PCR with anti-metazoan EUK565F_NGS/UnonMet_R primers) appeared comparable in terms of target sequences amplification efficiency, while the second protocol is faster and implies less loss of material. PCR protocols with microsporidia-specific primer systems 18F/530R and V1F/530R appeared equally suitable for amplification of a wide range Microsporidia corresponding to different phylogenetic lineages and also contributed to rozellids and aphelids dataset.

After data processing we obtained 56 OTUs (366 reads) of Microsporidia, 112 OTUs (954 reads) of Aphelida and 450 OTUs (3818 reads) of Rozellida. The discovered sequences contributed to molecular clades within the entire basal Holomycota tree, including both described species (*Rozella*, *Paraphelidium*, *Aphelidium*, *Paramicrosporidium*, *Mitosporidium*, *Morellospora* and numerous microsporidian species) and consistently reproduced in previous phylogenetic studies environmental clades (NAMAKO37, LKM46, NAMAKO35, hww6, LKM15, AZeuk2, LKM11, WIM27, dpeuk6, LKM46). Numerous sequences clustering with microsporidian sequences from aquatic habitats from clades 4(IV), 1(I), 3(V) [15, 16] were detected. A potentially new group of early microsporidia basal to Core Microsporidia + Metchnikovellida + RL107-1 (FN546176) was identified.

All analyzed samples showed rather similar diversity profiles, except the one taken from an especially silted area of Turovskoye Lake (T.07.21). Sequences contributed to most microsporidian clades, *Mitosporidium* clade, *Rozella* clade and Aphelida did not show any visible correspondence to sampling site or sampling time, probably following a continuous distribution pattern of potential multicellular hosts in closely located sampling sites. Sequences contributed to *Morellospora*, *Paramicrosporidium* and *Nucleophaga* clades shared patchy and sporadic distribution pattern of potential unicellular hosts sensitive to unstable microhabitat conditions.

Conclusion: Since a significant fraction of obtained microsporidian and rozellids sequences formed clades with environmental sequences from Genbank, and these clades lack any described representatives, phylogenetic diversity of basal Holomycota in different ecosystems clearly remain underdescribed and require further study.

Funding: The study was supported by the Russian Science Foundation (project No. 23-74-00071) and used equipment of the Core Facility Centers “Biobank”, “Development of Molecular and Cell Technologies” and “Culture Collection of Microorganisms” of the Research Park of Saint Petersburg University.

References

1. Karpov S., Mamkaeva M., Aleoshin V., Nassonova E., Lilje O., Gleason F. Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Front Microbiol.* 2014;5:112. doi 0.3389/fmicb.2014.00112
2. Lafferty K., Allesina S., Arim M., Briggs C. et al. Parasites in food webs: the ultimate missing links. *Ecol Lett.* 2008;11(6):533-46. doi 10.1111/j.1461-0248.2008.01174.x
3. Murareanu B., Sukhdeo R., Qu R., Jiang J., Reinke A. Generation of a Microsporidia species attribute database and analysis of the extensive ecological and phenotypic diversity of Microsporidia. *mBio* 2021;12(3):e0149021. doi 10.1128/mbio.01490-21
4. Bass D., Chech L., Williams B., Berney C., Dunthon M., Mahe F., Torruella G., Stentiford G., Williams T. Clarifying the relationships between microsporidia and cryptomycota. *J Eukaryot Microbiol.* 2018;65(6):773-782. doi 10.1111/jeu.12519
5. Chauvet M., Monjot A., Lepere C. High diversity of microsporidian parasites and new planktonic hosts in freshwater and marine ecosystems. *Limnol Oceanogr.* 2023;68:928-941. doi 10.1002/lno.12321

6. Carnegie R., Meyer G., Blackbourn J., Cochennec N., Franck B., Bower S. Molecular detection of the oyster parasite *Mikrocytos mackini*, and a preliminary phylogenetic analysis. *Dis Aquat Org.* 2003;54:219-27. doi 10.3354/dao054219
7. Taya Y., Kinoshita G., Mohamed W., Moustafa M., Ogata S., Chatanga E., Ohari Y., Kusakisako K., Matsuno K., Nonaka N., Nakao R. Applications of blocker nucleic acids and non-metazoan PCR improves the discovery of the eukaryotic microbiome in ticks. *Microorganisms.* 2021;9(5):1051. doi 10.3390/microorganisms9051051
8. Bass D., Del Campo J. Microeukaryotes in animal and plant microbiomes: Ecologies of disease? *Eur J Protistol.* 2020;76:125719. doi 10.1016/j.ejop.2020.125719
9. Weiss L., Vossbrinck C. Molecular biology, molecular phylogeny and molecular diagnostic approaches to the Microsporidia. In: Wittner M., Weiss L.M. The Microsporidia and Microsporidiosis. Amer Society for Microbiology, 1999. doi 10.1128/9781555818227.CH4
10. Schloss P. Reintroducing mothur: 10 years later. *Appl Environ Microbiol.* 2020;86:e02343-19. doi 10.1128/AEM.02343-19
11. Katoh K., Rozewicki J., Yamada K. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinf.* 2019;20(4):1160-1166. doi 10.1093/bib/bbx108
12. Gouy M., Guindon S., Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 2010;27(2):221-224. doi 10.1093/molbev/msp259
13. Nguyen L., Schmidt H., von Haeseler A., Minh B. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268-274. doi 10.1093/molbev/msu300
14. Letunic I., Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016;44(W1):W242-W245. doi 10.1093/nar/gkw290
15. Vossbrinck C., Debrunner-Vossbrinck B. Molecular phylogeny of the Microsporidia: ecological, ultrastructural and taxonomic considerations. *Folia Parasitol (Praha)* 2005;52(1-2):131-142. doi 10.14411/fp.2005.017
16. Vossbrinck C., Debrunner-Vossbrinck B., Weiss L. Phylogeny of the Microsporidia. In: Weiss L.M, Becnel J.J. (Eds.). Microsporidia: Pathogens of Opportunity. John Wiley & Sons, 2014;203-220. doi 10.1002/9781118395264.ch6

doi 10.18699/bgrs2024-5.2-19

Phylogenomics reveals early evolution of microsporidia

Nassonova E.[1]*, Rayko M.[1], Bondarenko N.[1], Frolova E.[1, 2], Kamyshatskaya O.[1], Smirnov A.[1, 2]

[1] *Laboratory of Cytology of Unicellular Organisms, Institute of Cytology, RAS, St. Petersburg, Russia*

[2] *Department of Invertebrate Zoology, Faculty of Biology, Saint Petersburg University, St. Petersburg, Russia*

* *nosema@mail.ru*

Key words: microsporidia; intracellular parasites; reductive evolution; phylogenomics

Motivation and Aim: Microsporidia are a widespread and species-rich group of obligate intracellular parasites of animals and some protists. They combine highly specialized and primitive features in their organization. The biosynthetic pathways and molecular machineries are seriously reduced in these parasites, and the genomes are small and compactly organized. Recent metagenomic studies demonstrated a huge "hidden" diversity of early microsporidia and related groups of organisms, while only a limited number of described representatives is known. The isolation and thorough study of new organisms from this group is of special interest. One of the main challenges is to reconstruct the early steps of evolutionary history of this remarkable group of parasites, which was strongly influenced by reductive evolution.

Methods and Algorithms: To get a sufficient amount of DNA for genome sequencing of miniature intracellular organisms, we applied a single-cell genomics approach. Whole genome amplifications of DNA from individually isolated infected host cells were performed using multiple displacement amplification (MDA). The MDA products were checked for the presence of target DNA using PCR with microsporidia-specific primers for amplification of the SSU rRNA gene. Verified MDA products were used for library preparations and sequencing on the Illumina HiSeq2500 platform. The genomes were assembled using SPAdes in a single-cell mode and binned with MaxBin2, to identify the target bin within the primary metagenomic assembly (parasite + host + bacteria). The resulting target bins were checked for completeness and contamination using BUSCO and BlobTools. The phylogenomic trees based on the BUSCO Fungi dataset were built using a set of publicly available genomes of microsporidia and related organisms. Augustus program was used for *ab initio* determination of functional sequences. Repetitive elements in the genomes were searched with RepeatModeler and the Dfam database. Functional annotation was carried out using the eggNOG-mapper package. Prediction of functional domains was based at the PFAM and eggNOG databases. Functional orthologs of the predicted proteins were determined using the KEGG database.

Results: Early microsporidia were isolated from a broad range of hosts: from protists (gregarines from marine annelids, free-living amoebae) to Metazoa (tardigrades, insects). Several new species were identified. The sequenced genomes were rather small and compact. They varied in size from 3.5 to 9.1 Mbp; from 2000 to 3750 ORFs were identified; gene density varied in the range from 0.45 to 0.65 gene/kb. The diversity of mobile elements was represented mainly by simple repeats, rolling circles and low

complexity elements, while in some lineages LINEs, LTR elements and DNA transposons were rather abundant. In some lineages we observed the complete or partial reduction of the mitochondrial genome, while in others, it was not affected by a tendency to degeneration. Phylogenomic analysis suggested the possible branching pattern of early microsporidian lineages. It also demonstrated the widespread occurrence of co-infections with two or more hyperparasitic microsporidian species in one superhost. For example, in the case of metchnikovellid parasites of gregarines from marine annelids, the co-occurring species (apparently represented only by the early proliferative stages) often remained hidden until genomic data became available. The comparative genomic analysis supposed that early evolution of microsporidia involved a series of independent losses and acquisitions of genes and molecular machineries in different lineages, probably as a result of adaptations to the various hosts.

Conclusion: Phylogenomic and comparative genomic analyses of known and recently isolated microsporidia allowed to propose a hypothetical scenario of the early evolution in this group.

Funding: The study is supported by the Russian Science Foundation (project No. 23-74-00071). This study utilized equipment of the Core Facility Centres “Biobank”, "Development of Molecular and Cell Technologies” and “Culture Collection of Microorganisms” of the Research Park of Saint Petersburg University.

doi 10.18699/bgrs2024-5.2-20

The significance of genomic studies of Aphelida (Aphelida, Hooycota, Opisthokonta) for solving questions of evolution in the Opisthokonta group

Pozdnyakov I.R.[1]*, Selyuk A.O.[2], Vishnyakov A.E.[2], Karpov S.A.[1, 2]

[1] *Zoological Institute, Russian Academy of Sciences, St. Petersburg, Russia*

[2] *Department of Invertebrate Zoology, Faculty of Biology, St. Petersburg University, St. Petersburg, Russia*

* *d_igor_po@yahool.com*

Key words: genomic studies; Aphelida; Hooycota; Opisthokonta

Motivation and Aim: Aphelids [1, 2] are the significant object for evolutionary research because:
1) They are the closest evolutionary relatives of Fungi, but do not have morpho-physiological characteristics specific to fungi (osmotrophy, mycelial structure with apical growth, etc.);
2) They have a complex life cycle with alternating immobile plasmodial and motile amoebo-flagellate stages. This cycle, on the one hand, coincides with the life cycles of zoosporic fungi and unicellular Holozoa; on the other hand, there is primary data on the differences of the genetic bases of these life cycles [3, 4].
Thus, genomic and proteomic studies of aphelids allow:
1) Explore the prerequisites and initial stages of the fungal evolution.
2) Raise and solve questions on the evolution of cellular polymorphism in unicellular Opisthokonta and its connection with the origin of multicellularity in filamentous fungi and Metazoa.
We have performed:
1. Assessment of the possibility of enhancing the osmotrophic abilities in aphelids by analyzing the quantitative and qualitative composition of their transmembrane transporter proteins of the MFS (Major facilitator superfamily) superfamily [5, 6].
2. Study of gene expression in *Aphelidium insulamus* at the plasmodium stage.
Methods and Algorithms:
For analysis of MFS proteins:
Major facilitator superfamilyThe MFS-domain proteins initially were selected based on annotations among the hits of the BLASTP searching in the predicted proteomes of selected organisms with annotated genomes. The *Saccharomyces cerevisiae* proteins were taken as the initial queries. The hidden Markov models were built by the hmmbuild program of the hmmer, v.3.3.2 batch [7] for necessary protein families. The final search was performed using hmmsearch among predicted proteoms of representatives of various subgroups of Opisthokonta including four aphelid species (*Amoeboaphelidium protococcarum*, *Amoeboaphelidium occidentale*, *Paraphelidiun tribonematis* and *Aphelidium insulamus*). The multiple sequence alignment (MSA) was prepared in the M-Coffee aligner using the web server interface (https://tcoffee.org). The MSA was treated in TrimAl, v.1.4.rev15 with a gap threshold of 0.5. For the tree construction was

used IQ-Tree 2, v.2.0.3 [8] with determined LG+F+G4 substitution model and 100000 replicates of ultrafast bootstrap.
For gene expression analysis:
Library preparation for RNA-seq was performed using the NEBNext® Single Cell/Low Input RNA Library Prep Kit from New England Biolab (NEB) using a protocol for low input. Sequencing was performed on an Illumina HiSeq sequencer. For each sample, 15 million short paired reads 2x100 were obtained. Short reads were aligned to the already known A. insulamus transcriptome and quantified using the utitlites included in the BBtools package [9]. Further quantified reads were analysed in DEseq2 [10].
Results and Conclusions:
1. It has been shown that studied aphelids lack MFS proteins specific for fungi (both a specific fungal protein family Drug:H+ antiporters-2 (DAH-2) and specific fungal orthologs of the common sugar porters (SP) family) (Fig. 1). The repertoire of SP orthologs in aphelids turned out to be less diverse than in free-living opisthokonts, and one of the most limited among opisthokonts.
We argue that aphelids do not show signs of similarity with fungi in terms of their osmotrophic abilities. Moreover, the osmotrophic abilities of aphelids appear to be reduced in comparison with free-living unicellular opisthokonts.
Thus, it was concluded that the specific evolution of fungi began after the separation of the fungal and aphelid lineages.
2. Groups of genes with the similar expression level, genes having an expression level an order of magnitude greater than the median and genes with the maximum expression level in *A. insulamus* plasmodia were identified. Assumptions are made about the most important metabolic processes for plasmodia and genes whose products are potentially involved in interaction with the host and in the regulation of switching stages of the life cycle are noted.
It is planned to obtain gene expression profiles at other stages of the life cycle and build a general picture of changes in gene expression, metabolic and regulatory processes. Further inclusion of the obtained data in comparative studies.
Funding: The study is supported by the Russian Science Foundation grant No. 21-74-20089.

References

1. Karpov S.A., Mamkaeva M.A., Aleoshin V.V. et al. Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Front Microbiol*. 2014;5:112. doi 10.3389/fmicb.2014.00112
2. Galindo L.J., Torruella G.; López-García P. et al. Phylogenomics, supports the monophyly of aphelids and fungi and identifies new molecular synapomorphies. *Syst Biol*. 2023;72(3):505-515. doi 10.1093/sysbio/syac054
3. Torruella G., de Mendoza A., Grau-Bové X. et al. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Curr Biol*. 2015;25(18):2404-2410. doi 10.1016/j.cub.2015.07.053
4. Pozdnyakov I.R., Zolotarev A.V., Karpov S.A. Comparative analysis of zoosporogenesis' genes of the bastoclad Blasto-cladiella emersonii and the aphelid Paraphelidium tribonematis reveals the new directions of evolutionary research. *Protistology*. 2021;15(1):10-23. doi 10.21685/1680-0826-2021-15-1-2
5. Pao S.S., Paulsen I.T., Saier M.H. Jr. Major Facilitator Superfamily. *Microbiol Mol Biol Rev*. 1998;62(1):1092-2172. doi 10.1128/MMBR.62.1.1-34.1998
6. Gonçalves C., Coelho M.A., Salema-Oom M., Gonçalves P. Stepwise Functional Evolution in a Fungal Sugar Transporter Family. *Mol Biol Evol.* 2016;33(2):352-366. doi 10.1093/molbev/msv220
7. Wheeler T.J., Eddy S.R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics*. 2013;29(19):2487-2489. doi 10.1093/bioinformatics/btt403

8. Nguyen L.T., Schmidt H.A., von Haeseler A., Minh B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268-274. doi 10.1093/molbev/msu300
9. Bushnell B., Rood J., Singer E. BBMerge – Accurate paired shotgun read merging via overlap. *PLoS One*. 2017;12(10):e0185056.
10. Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi 10.1186/s13059-014-0550-8

doi 10.18699/bgrs2024-5.2-21

An approach to assembling MAGs of intracellular parasites with a reduced genomes from MDA data

Rayko M.*, Nassonova E.
Laboratory of Cytology of Unicellular Organisms, Institute of Cytology RAS, St. Petersburg, Russia
** mike.rayko@gmail.com*

Key words: MAG; metagenomics; microsporidia; intracellular parasites

Motivation and Aim: Representatives of the group Opisthophagea are of interest from the evolutionary point of view. This group comprises two large lineages, Rozellida (also known as Rozellosporidia) and Microsporidia. In phylogenetic trees, they form a separate branch of Holomycota, closely related to the "true" fungi. Microsporidia comprise of a great diversity of species, whereas few representatives have been described for Rozellida. However, recent metagenomic studies show that both these lineages are abundant and inhabit diverse environments. They are intracellular parasites (in many cases even hyperparasites), and they show genome reduction, when part of their functions is performed at the expense of the host genome [1].
Their study is complicated by the fact that they are very difficult to isolate in sufficient numbers for sequencing. In most cases, we deal with metagenomes, and the MDA method is used to obtain material. This method gives uneven coverage, which further complicates sequencing and data analysis. As a result, metagenome-assembled genomes are incomplete, and we need a way to distinguish between genome elements lost during parasitic reduction, or simply not assembled due to various reasons – small amounts of DNA, low coverage, or complex community composition [2].
Methods and Algorithms: We assembled a pipeline to analyse MDA data from genome sequencing of different Opisthophagea specimens. The metagenome is assembled using SPAdes in a single-cell mode (adapted for uneven coverage), then binning with MaxBin2, the resulting bins are checked for completeness and contamination using BUSCO and BlobTools. Finally, we identify the target bin (MAG), and build a phylogenomic tree based on the BUSCO Fungi dataset using a set of publicly available microsporidian genomes. We use both a supermatrix-based approach (general concatenated alignment and phylogenetic reconstruction using the maximum likelihood method) and a supertree-based method – building a separate tree for each of the orthologs and obtaining a tree using Astral.
The resulting tree allows us to determine the taxonomic position of the studied specimen. We then distinguish the clade which the specimen belongs to, using branch length threshold, and then build a loss pattern for each of the genes of interest. We assume that the loss of a gene during reduction must be a synapomorphy, and that the absence of a gene for technical reasons may occur randomly and not be consistent with the topology of the tree. Thus for a range of genes we can reliably detect parasitic loss events.
Results: We analyzed several dozen samples of MDA metagenomes containing microsporidia genomes using our pipeline. For each sample, a target bin was selected, followed by further decontamination and quality assessment of the resulting MAG. The completeness assessment values by BUSCO for microsporidia were in range 20-45 %, and these values were consistent across all samples and were similar within clades. It is

also important to note that for publicly available genomes we obtained similar values. Thus, these values reflect parasitic genome reduction. By using phylogenomic analysis results, we compared the tree topologies for each ortholog with a consensus final tree and identified genes with shared patterns of losses.

Conclusion: The application of the developed approach and accumulation of information allows for the generation of a sufficiently accurate phylogenetic tree for various branches of Opisthophagea, as well as tracking gene loss patterns during parasitic genome reduction. With further increase in the number of specimens, the picture of the evolution of the entire Holomycota branch will become much clearer.

Funding: The study is supported by the Russian Science Foundation (project No. 23-74-00071).

References

1. Frolova E.V., Paskerova G.G., Smirnov A.V., Nassonova E.S. Diversity, distribution, and development of hyperparasitic microsporidia in gregarines within one super-host. *Microorganisms.* 2023;11:152. doi 10.3390/microorganisms11010152
2. Nassonova E.S., Bondarenko N.I., Paskerova G.G. et al. Evolutionary relationships of *Metchnikovella dogieli* Paskerova et al., 2016 (Microsporidia: Metchnikovellidae) revealed by multigene phylogenetic analysis. *Parasitol Res.* 2021;120(2):525-534. doi 10.1007/s00436-020-06976-x

doi 10.18699/bgrs2024-5.2-22

Phylogenetic analysis and molecular dynamic simulations of CsqR, a transcription factor from *Escherichia coli*

Rybina A.A.[1]*, Glushak R.A.[2], Bessonova T.A.[3, 5], Dakhnovets A.I.[1], Rudenko A.Y.[4], Ozhiganov R.M.[4], Kaznadzey A.D.[5], Tutukina M.N.[1, 3, 5], Gelfand M.S.[1]

[1] *Skolkovo Institute of Science and Technology, Moscow, Russia*

[2] *Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia*

[3] *Institute of Cell Biophysics, Federal Research Center "Pushchino Scientific Center for Biological Research RAS", Pushchino, Russia*

[4] *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

[5] *Institute for Information Transmission Problems, RAS, Moscow, Russia*

* *rybinaann@gmail.com*

Key words: phylogeny; sulfoquinovose; molecular docking; molecular dynamic simulations

Motivation and Aim: CsqR is a local transcription factor from *Escherichia coli* that regulates expression of *yih* genes responsible for the degradation of sulfosugar sulfoquinovose (SQ) [1, 2]. We recently demonstrated that lactose could potentially regulate the *yih* genes [2]. SQ and its derivatives sulfoquinovosyl glycerol (SQG) and sulforhamnose (SR) were reported to act as putative effectors of CsqR [3]. Preliminary experimental findings from our laboratory suggested the presence of an alternative truncated variant of CsqR. Here, we aimed to study the evolution of CsqR and estimate patterns of its binding to potential effectors SQ, SR, SQG, and lactose.

Methods and Algorithms: Phylogenetic tree of CsqR protein homologs was constructed to study evolutionary patterns of CsqR. Production of the CsqR protein forms was validated in the western-blot experiment. To model interactions of CsqR with candidate effectors, predicted protein structures of CsqR were optimized in 2 μs molecular dynamic simulations and then subjected to blind molecular docking with potential ligands SQ, SR, SQG, and lactose.

Results: Phylogenetic analysis revealed the presence of CsqR homologs in both the Actinobacteria and Proteobacteria phyla. The structure of the tree suggested that *csqR* underwent duplication at some point. Notably, CsqR homologs of Enterobacteriales species had a highly conserved Met25 residue. We suggested the existence of two alternative variants of CsqR: a full-length version (CsqR-l) and a shorter variant lacking 24 N-terminal residues (CsqR-s). Western blot analysis showed that CsqR was produced in two protein forms, 28.5 kDa (CsqR-l) and 26 kDa (CsqR-s), the latter starting translation at Met25. Interestingly, CsqR-s exhibited significant activation during growth with sulfoquinovose as the sole carbon source, displacing CsqR-l during the stationary phase of growth on a rich medium. According to molecular dynamic simulations, CsqR-s might have two potential structural arrangements, with the interdomain linker appearing either as a disordered loop or an α-helix. This helix allowed the N-terminal domain to rotate in a hinge-like motion, resulting in the transition of CsqR-s between two conformations referred to as "open" and "compact". Molecular docking suggests that CsqR-s may have a greater ability to discriminate between putative ligands and other compounds compared to CsqR-l, indicating potential differences in ligand binding affinity or specificity between the two variants

Conclusion: Bacteria rarely possess multiple forms of a single protein, with only only a handful of cases reported to date. CsqR might be an interesting inclusion in this group, poised for additional experimental investigation.
Funding: The study is supported by the RSF via grant 24-14-00276.

References

1. Denger K. et al. Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature.* 2014;507(7490):114-117. doi 10.1038/nature12947
2. Kaznadzey A. et al. The genes of the sulphoquinovose catabolism in *Escherichia coli* are also associated with a previously unknown pathway of lactose degradation. *Sci Rep*. 2018;8(1):3177. doi 10.1038/s41598-018-21534-3
3. Shimada T. et al. Regulatory role of CsqR (YihW) in transcription of the genes for catabolism of the anionic sugar sulfoquinovose (SQ) in *Escherichia coli* K-12. *Microbiology.* 2019;165:78-89. doi 10.1099/mic.0.000740x

doi 10.18699/bgrs2024-5.2-23

Analysis of SARS-CoV-2 genomic data using haplotype networks

Samoilov A.E.*, Chudinov I.K.
Research Institute for Systems Biology and Medicine, Moscow, Russia
** andrei.samoilov@gmail.com*

Key words: SARS-CoV-2; haplotype networks; viral surveillance

Motivation and Aim: Since the beginning of the COVID-19 pandemic, the world scientific community has faced the problem of analyzing SARS-CoV-2 genomic sequences. One of the methods of visualizing and analyzing genetic information is the construction of haplotype networks, which allow overcoming the limitations of classical methods of phylogenetics. Here we have utilized haplotype networks to detect and characterize imports of SARS-CoV-2 to Russia.
Methods and Algorithms: Genomic sequences of SARS-CoV-2 longer than 29,000 bp and N content lower than 1 % belonging to pangolin [1] lineages BA.1, BA.2 (excluding BA.2.75, XBB and later lineages) and Delta were selected from GISAID [2] and VGARus databases. Filtration was performed with Nextclade [3] and R script. Missing data in non-full Russian genomes was restored using a python script. Haplotype network graphs were built and visualized using R script.
Results: Filtration of genomes was performed as follows: for each lineage among the analyzed (BA.1, BA.2 and Delta), the list of the parent lineage substitutions and deletion was obtained by analyzing the most frequent mutations. Also the list of true reversion substitutions (i.e. mutation occurring at the same position for the second time, effectively reverting the nucleotide to the one present in the Wuhan genome) associated with certain pangolin lineages was created. Genomes with pseudoreversions (or reversion mutations not associated with a certain lineage and most likely caused by bioinformatic error) were filtered out, as well as low quality genomes.
After the filtration, to increase the amount of Russian genomes in the analysis, genomes with N nucleotides were repaired by aligning them against closest full genome sequences collected earlier and replacing Ns with consensus sequence in the N regions or discarded if there was no consensus in the alignment in the N regions.
To create haplotype network graphs, we have determined all of the potential ancestors and descendants of the analyzed Russian genomes (a substitution mutations of a potential ancestor genomes are a subset of a substitution mutations of a potential descendant genome) and created a directed graph (with the direction from the root to the increasing amount of substitutions, in accordance with the direction of the evolution) by connecting each observed substitution pattern to all of its potential ancestors, excluding the potential ancestors of potential ancestors. An example of haplotype network graph for Delta lieage is presented on the Figure 1. Next, the obtained haplotype network graph was trimmed by selecting more likely ancestor in cases where there was more than one using following algorithm: firstly, the closest ancestor was selected, secondly, the ancestor detected in the same country or countries as the descendant was selected, lastly, the earliest ancestor was selected. An example of haplotype network graph for Delta lieage is presented on the Fig. 1. After trimming, we were able to obtain imports of SARS-CoV-2 into Russia

by obtaining every substitution pattern found in Russia and whose closest ancestor was not detected in Russia.

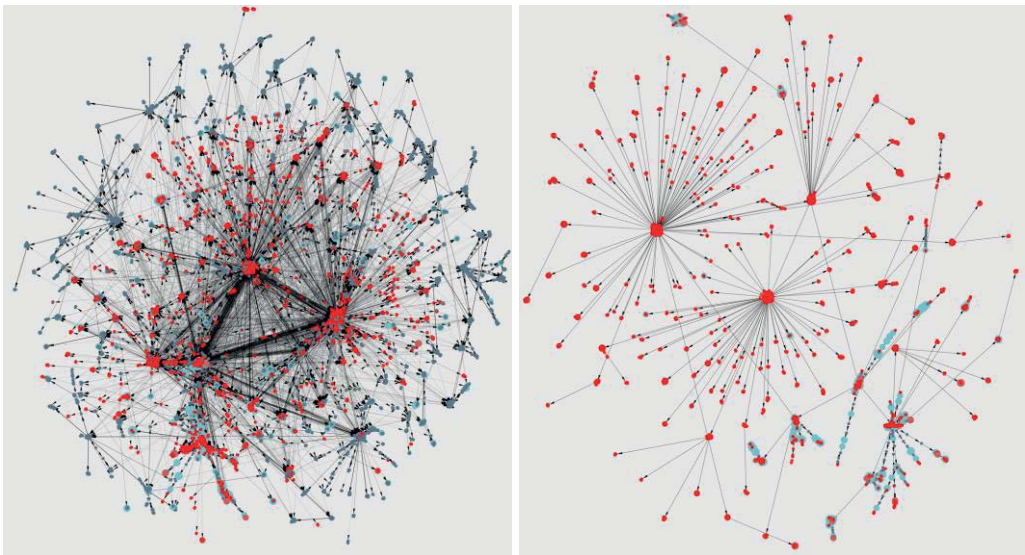


Fig. 1. Haplotype networks of Russian SARS-CoV-2 genomes. Graph vertices represent substitution patterns (a certain combination of substitution mutations), which have been observed in SARS-CoV-2 genomes at least once. Direction of the haplotype network graph represents an increase in the number of mutations from the root (parent genome for the lineage) to the increased amount of substitutions. Russian substitution patterns (detected only in Russia or detected in Russia first) are colored in red, substitution patterns detected in single country outside Russia are colored in dark grey, patterns detected in several countries outside Russia are colored in blue. The size of the vertices is increased with the amount of countries this substitution pattern was observed in. Trimmed haplotype networks are created from untrimmed ones by selecting the most likely path to the root. Left: untrimmed haplotype network graph of Delta lineage in Russia. Right: trimmed haplotype network graph of Delta lineage in Russia

We have detected 1012 imports of Delta lineage, 773 imports of BA.1 lineage and 569 imports of BA.2 to Russia. Since Delta existed for about 10 months, while BA.1 and eary BA.2 existed for a few months each, the low abount of Delta imports suggests that anti-epidemic measures which were applied in 2021 were effective way to decrease the number of imports of SARS-CoV-2 in Russia. Another interesting result is the comparison of our data with the paper by Klink et al. [4], which suggests that 90 % or Russian Delta genomes were descendants of one single import. Our data shows that only 60 % of Russian Delta genomes were descendants of that import event, while the other 30 % were re-imported repeatedly over 500 times during all of the course of pandemic.
Conclusion: Haplotype networks are an efficient method of analysis, which allows to visualize and characterize big genomic data during massive pandemic events.
Funding: The study is supported by the subsidy of the Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing No. 141-02-2023-208.

References

1. O'Toole Á. et al. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. *BMC Genomics*. 2022;23(1):121. doi 10.1186/s12864-022-08358-2
2. Khare S. et al. GISAID's role in pandemic response. *China CDC Weekly*. 2021;3(49):1049. doi 10.46234/ccdcw2021.255
3. Aksamentov I., Roemer C., Hodcroft E.B., Neher R.A. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J Open Source Softw*. 2021;6(67):3773. doi 10.21105/joss.03773
4. Klink G.V., Safina K.R., Nabieva E., Shvyrev N., Garushyants S., Alekseeva E. et al. The rise and spread of the SARS-CoV-2 AY. 122 lineage in Russia. *Virus Evol*. 2022;8(1):veac017. doi 10.1093/ve/veac017

doi 10.18699/bgrs2024-5.2-24

New classification of prokaryotic DNA methyltransferases

Samokhina M.[1, 2]*, Alexeevski A.[1, 2, 3]

[1] *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

[2] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

[3] *Research Institute for System Analysis, RAS, Moscow, Russia*

* *mashila6799@gmail.com*

Key words: DNA methyltransferases; classification; evolution

Motivation and Aim: Prokaryotic DNA methyltransferases (MTases) are essential enzymes in bacterial immune systems like restriction and modification systems and others. Also the enzymes play an important role in regulation gene expression, cell cycle and DNA repair. MTase classification by sequence similarity was built more than 30 years ago [1]. Our goal was to build new MTases classification that describe modern diversity of MTases and construct an evolutionary model.

Methods and Algorithms: To built classification we use MTase sequences from database REBASE. For splitting MTases into groups we used HMM profiles from Pfam and SUPFAM databases. Profile hits were used to classify MTases into classes.

Results: We divided all MTases into 10 classes based on profile region hits (Fig. 1). The classes are named in Latin letters from A to J. Each class has a specific topology. Topologies A–D obtained from data on resolved spatial structures, the remaining topologies are obtained from the predicted structures from AlphaFold database. We identified catalytic motifs in all MTases. Based on the similarity of catalytic motifs, MTases are divided into 3 evolutionary groups. Within each group, the sequences of catalytic motifs intersect, but between groups they do not intersect.

We have developed an automatic algorithm that detects the topology of MTase by region hits of cat-profiles in MTase sequences (https://github.com/MVolobueva/MTase-classification.git) and a web interface to run it (https://mtase-pipeline-6g1yfq9ugw8.streamlit.app/) [2].

We have compared the developed classification with existing ones, namely into methylation types, restriction-modification system (R-M) types they belong and permutations of S-Adenosyl methionine (SAM)-binding motifs, calalytic motifs and TRD (target recognition djvfin) (Table 1). MTases from class C and K have m5C methylation type. MTases from the P-M I system are found only in class A, MTases from the P-M III system are found only in class B.

The comparison with the classification by motifs shows that classes A, L, K, F correspond to class γ, classes M and D correspond to class α. The K class was previously allocated to a separate class ζ*. However, there is no experimental evidence that class ζ is different from class γ.

Conclusion: We have built new MTase classification based on sequence similarity of catalytic domain. All known MTases felt into 10 classes that comprise 3 evolutionary groups.

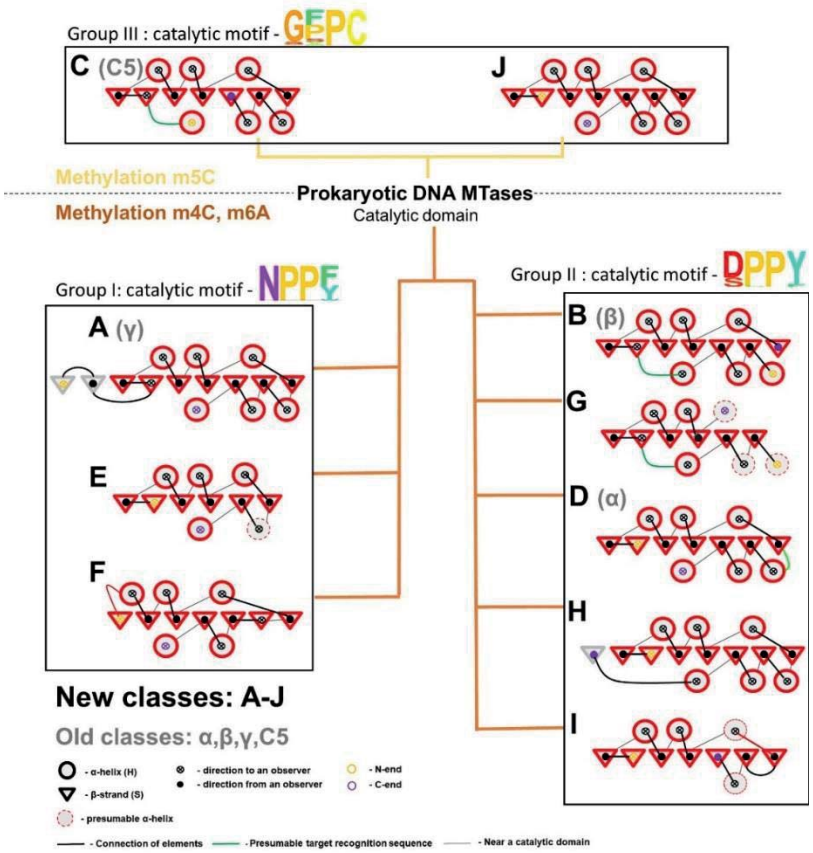


Fig. 1. New classification of MTases. Classes are named in Latin letters from A to J. For each class, a topology scheme is provided, which displays the sequence of connection of the secondary structure elements

Table 1. Comparison of the new classification with the previous classifications

Class	Methylation type	R-M type	Subtype
A	m4C, m6A	I, II	γ
B	m4C, m6A	II, III	β
D	m4C, m6A	II	α
L	m4C, m6A	II	γ
M	m4C, m6A	II	α
C	m5C	II	C5
K	m5C	II	ζ*, γ
E	m6A	II	
F	m6A	II	γ
G	m6A	II	
I	m4C, m6A	II	γ

* Buinitsky [1] classified zeta into a separate class.

References

1. Bujnicki J.M. Sequence permutations in the molecular evolution of DNA methyltransferases. *BMC Evol Biol*. 2002;2:3. doi 10.1186/1471-2148-2-3
2. Samokhina M., Rusinov I., Alexeevski A. Classification of prokaryotic DNA methyltransferases by topology and sequence similarity. *biorXiv*. 2023. doi 10.1101/2023.12.13.571470

doi 10.18699/bgrs2024-5.2-25

Evolution and distribution of prokaryotic type III toxin-antitoxin systems

Selifonov I.[1, 2]*, Moshenskiy D.[2], Alekseevskiy A.[1, 2, 3]

[1] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

[2] *Department of mathematical methods in biology, A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

[3] *Research Institute for System Analysis, RAS, Moscow, Russia*

* *selifonov2002@gmail.com*

Key words: bacteria; defense systems; toxin-antitoxin systems

Motivation and Aim: A prokaryotic toxin-antitoxin (TA) system consists of two main components: a toxin that inhibits growth or triggers programmed cell death and a corresponding less stable antitoxin that neutralizes the former. Though these systems can be considered as selfish genetic elements that enforce their own stable vertical inheritance, for many of them a wide variety of functions have been proposed [1]. Type III TA systems consist of a protein toxin and a small ncRNA antitoxin, which is encoded in a series of repeats directly preceding the toxin gene. The latest thorough bioinformatic analysis of the evolution and distribution of these systems was performed by Blower et al. in 2012, when they were separated into three families: toxIN, tenpIN and cptIN based on the sequence similarity of the toxins [2]. In this work we revisit this topic with new genomic data. The aim of this work is to assess the diversity and phylogenetic distribution of prokaryotic type III TA systems.

Methods and Algorithms: Using the 125 classified toxin sequences, we built multiple sequence alignments for each of the 3 families. After manual curation and trimming of the alignments they were used to create 3 HMM profiles. Using the profiles, we performed searches in the original protein sequences, which let us define the bit score gathering thresholds for the 3 toxin groups. For each protein family the value of the threshold was set to the score of the weakest (the least statistically significant) correct hit of this family's profile. Though the real score distribution of type III toxins is likely not as narrow as the one observed in the sample of 125 proteins, the use of such strict thresholds helped reduce the amount of noise in the data, which could otherwise complicate its further analysis.

The HMM profiles were then used to search for toxin homologs in 40359 prokaryotic genomes downloaded from the NCBI database. The resulting protein sequences with domain scores equal to or higher than the thresholds were included in the corresponding families. In order to identify the antitoxin DNA repeats we searched for sequence motifs in the fragments including 800 bases upstream of the coding sequence of each toxin gene.

To characterize the evolutionary dynamics of type III TA systems phylogenetic trees were built for unique toxin sequences and compared with reference trees created for the same bacteria based on the protein sequences for 107 essential single-copy core genes.

Results: By the described approach 610 putative toxins were found and classified (Table 1). For 484 of them we were also able to identify their cognate antitoxins.

Table 1. The number of found toxins for each family and the results of searching for antitoxin repeats

Family	Total systems	Number of repeats	Repeat length	No repeats found
toxIN	342	3.29 ± 2.07	33	33
cptIN	119	2.76 ± 1.90	39	37
tenpIN	149	2.37 ± 2.19	39	56

The newly described type III TA systems displayed an uneven phylogenetic distribution. Moreover, different taxa were enriched with toxins belonging to different families (Fig. 1).

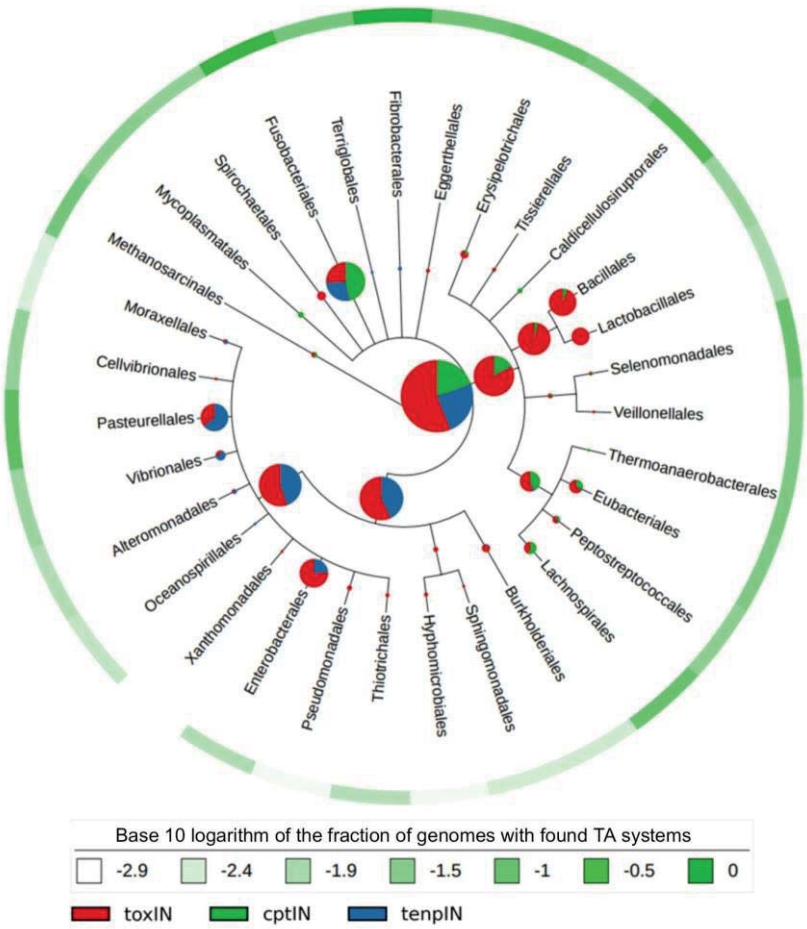


Fig. 1. The taxonomic distribution of type III TA systems of different families

The comparison of the phylogenetic trees built for unique toxin sequences and the corresponding strains of bacteria showed that TA loci frequently undergo horizontal gene transfer between different species and genera of microorganisms.

Conclusion: The 3 TA families differ in the sequence and organization of antitoxin repeats and are usually found in different taxa. Horizontal gene transfer between both closely related and distant groups of prokaryotes plays a significant role in the evolution and spread of type III TA systems.

Funding: The study is supported by the grant RSF No. 21-14-00135.

References

1. Jurėnas D., Fraikin N., Goormaghtigh F., van Melderen L. Biology and evolution of bacterial toxin–antitoxin systems. *Nat Rev Microbiol*. 2022;20(6):335-350. doi 10.1038/s41579-021-00661-1
2. Blower T.R., Short F.L., Rao F., Mizuguchi K., Pei X.Y., Fineran P.C., Luisi B.F., Salmond G.P.C. Identification and classification of bacterial Type III toxin–antitoxin systems encoded in chromosomal and plasmid genomes. *Nucleic Acids Res*. 2012;40(13):6158-6173. doi 10.1093/nar/gks231

doi 10.18699/bgrs2024-5.2-26

Single-cell genomics – a powerful tool to resolve the phylogeny of uncultivable species of Amoebozoa

Smirnov A.[1]*, Mesentsev E.[1], Kamyshatskaya O.[1, 2], Glotova A.[1, 2], Bondarenko N.[1], Rayko M.[2], Skalon E.[1], Lotonin K.[1], Nassonova E.[1, 2]

[1] *Department of Invertebrate Zoology, Faculty of Biology, Saint Petersburg University, St. Petersburg, Russia*

[2] *Laboratory of Cytology of Unicellular Organisms, Institute of Cytology, RAS, St. Petersburg, Russia*

** alexey.smirnov@spbu.ru*

Key words: Amoebozoa; Tubulinea; Discosea; Variosea; genomics; phylogeny; sequencing

Motivation and Aim: Amoebozoa still remain one of the least studied lineages of protists. Unstable body shape, difficulty of cultivation and study and inability of most species to grow in culture make investigation of the diversity and phylogeny of this group very difficult, and in some cases – almost impossible. Until recently, uncultivable species of Amoebozoa were absent in phylogenetic trees an their position in the system of Amoebozoa was defined based on evidences from morphology [1]. This problem seriously embarrassed further development of amoebozoan phylogeny [2]. Many species cannot multiply in laboratory culture and are recovered only in mixed, initial cultures and in a little number of specimens. To solve the problem, several years ago we elaborated and successfully applied protocols and approaches allowing us to extract DNA from a single amoebozoan cell and to perform the whole-genome amplification of this DNA [3, 4]. Using these techniques, we successfully found home for several poorly studied lineages of Amoebozoa using both single-gene phylogeny and phylogenomic approaches. In particular, in this way we have studied a very rare species *Thecochaos fibrillosum*, not seen by investigators for almost hundred years and established its position in the tree of Amoebozoa. We complemented the amoebozoan tree with numerous species belonging to the genus *Polychaos,* previously unavailable for molecular studies and made several more interesting findings in the filed of Amoebozoa phylogeny and phylogenomics.

Methods and Algorithms: For DNA isolation, individual amoeba cells were collected from the mixed cultures manually, with tapered-tip glass Pasteur pipette, or using Eppendorf micromanipulators/micronjector mounted on Leica DMI3000 inverted microscope equipped with IMC contrast. Individual cells were transferred into 40 mm Petri dishes filled with fresh Millipore-filtered (0.22 μm) PJ medium, and left to starve for three days. Every day the medium was replaced with the fresh one. Dishes with cells were examined using daily to check for the absence of visible food vacuoles inside cells, absence of fungal contamination, and other eukaryotes in the surrounding medium. Further, individual cells were collected and placed with 1–2 μl of the medium in 200-μl PCR tubes. DNA was extracted using the Arcturus PicoPure DNA Extraction Kit (Thermo Fischer Scientific, USA). We performed the whole genome amplification of the total DNA using Multiple Displacement Amplification (MDA) with slightly modified protocol for REPLI-g Single Cell DNA Amplification Kit (Qiagen, Hilden, Germany). The resulting MDA products were sequenced using Illumina HiSeq 2500 system. Quality control check of raw sequence data was performed using FastQC. SPAdes

assembler was used for *de novo* genome assembly [5]. ML analysis was performed by IQ-TREE v1.6.12 [6] with the model GTR+I+G+F. For multigene phylogeny, putative proteins were predicated using Metaeuk or TransDecoder (https://github.com/TransDecoder/TransDecoder/wiki). The base for the alignment was the dataset obtained by Kang et al. [7] (323 genes). Potential orthologs were identified using BLAST on protein datasets from potential closest relatives of studied organisms. ML trees were constructed for each protein-coding gene separately using FastTree [8] and examined manually to estimate the overall tree configuration and the potential presence of paralogs of the target gene. The set of positions for the analysis was selected using g-blocks as implemented in SeaView 4.0 [9]. RaxML 8.2.12 [10], and Phylobayes 1.8 [11] were run at Cipres portal [12].

Results: The above-described approach allowed us to perform several case studies and characterize a number of amoebozoan species, seen only in mixed culture and in very limited number of specimens. Until recently correct identification and record of such organisms was believed to be impossible. One of them is a member of the genus *Thecochaos* – the species *T. fibrillosum.* It was isolated from the soil of West Siberia, and we had ca. 15 cells of this organism. However, using single-cell techniques, we were able to obtain light- and electron-microscopic data, get SSU sequence and perform whole-genome amplification for NGS sequencing and build multigene phylogeny. The results of this study allowed this organism to find its home within the family Thecamoebidae (Amoebozoa, Discosea). The second case study is dedicated to very unusual amoeba with characteristic rolling movement of the cell surface. This organism was frequently seen from many habitats, but all attempts to culture it or to accumulate sufficient number of cells for DNA extraction failed. The organism is small (usually, 10–20 microns), thus single cells were collected using micromanipulation technique. This way to collect cells resulted in very clean DNA preparations. To establish its position in the phylogenetic tree of Amoebozoa we build a comprehensive multigene alignment consisting of ca. 70 000 AA positions. The resulting tree placed the studied organism in the Cutosea clade, which is congruent with LM and TEM data. The resulting Amoebozoan tree supported our hypothesis on the basal position of Tubulinea lineage in the tree of Amoebozoa [13]. In the latest studies we started to use not the whole cell, but the isolated cell nucleus to get DNA sample for the whole genome amplification. It opened a way to get sequences of largest species of Amoebida. These organisms are always filled with food vacuoles and endocytobionts. Until recently almost all attempts to get nuclear DNA from such species failed. Isolation of the nucleus allows getting pure genomic DNA and include these organisms in the phylogenetic and phylogenomic analysis [14].

Conclusion: Single-cell genomics is a powerful tool allowing phylogenetic placement of uncultivable amoeboid organisms found in the environment. This approach is especially fruitful in amoebozoan ecology, since identification of as many morphospecies as possible may be decisive for the quality of the study. Modern algorithms of data sorting and assemble allow to properly clear the target genome from contaminating reads. Using single-cell techniques, we were able to obtain light- and electron-microscopic data, get SSU sequence and perform whole-genome amplification for NGS sequencing and multigene phylogeny for several amoebae species. The number of cells necessary for a comprehensive study was reduced with these approaches from thousands to ten or less. Hence, establishing of stable culture is no more a necessary requirement for identification and proper description of amoeba species. Obtaining whole-genome data

allows direct building of phylogenomic trees, which are much more reliable rather than the single gene tree. These studies solved several old and frequently mentioned puzzles in Amoebozoan phylogeny and significantly improved the overall tree of Amoebozoa.

Funding: The study was supported by the Russian Science Foundation project 24-44-00096 and used equipment of the core facility centers “Culturing of microorganisms”, “Development of molecular and cell technologies” and “Biobank” of the Research Park of Saint Petersburg University.

References

1. Smirnov A., Chao E., Nassonova E., Cavalier-Smith T. A revised classification of naked lobose amoebae (Amoebozoa: Lobosa). *Protist.* 2011;162:545-570. doi 10.1016/j.protis.2011.04.004
2. Adl S.M., Bass D., Lane C.E., Lukeš J., Schoch C.L., Smirnov A. et al. Revisions to the classification, nomenclature, and diversity of Eukaryotes. *J Eukaryot Microbiol*. 2018;66:4-119. doi 10.1111/jeu.12691
3. Mesentsev Y., Bondarenko N., Kamyshatskaya O. et al. Thecochaos is not a myth: study of the genus Thecochaos (Amoebozoa, Discosea) – a rediscovered group of lobose amoeba, with short SSU gene. *Org Divers Evol.* 2023;23:7-24. doi 10.1007/s13127-022-00581-9
4. Kamyshatskaya O., Nassonova E., Smirnov A. Phylogenetic position of amoeba genus *Deuteramoeba* (Amoebozoa, Tubulinea). *Protistology.* 2017;11:231-237. doi 10.21685/1680-0826-2017-11-4-4
5. Nurk S., Bankevich A., Antipov D. et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol*. 2013;20:714-737. doi 10.1089/cmb.2013.0084
6. Nguyen L., Schmidt H., von Haeseler A., Minh B. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268-274. doi 10.1093/molbev/msu300
7. Kang S., Tice A.K., Spiegel F.W., Silberman J.D. et al. Between a pod and a hard test: the deep evolution of amoebae. *Mol Biol Evol.* 2017;34:2258-2270. doi 10.1093/molbev/msx162
8. Price M.N., Dehal P.S., Arkin A.P. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol.* 2009;26:1641-1650. doi 10.1093/molbev/msp077
9. Gouy M., Guindon S., Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 2010;27(2):221-224. doi 10.1093/molbev/msp259
10. Stamatakis A. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312-1323
11. Lartillot N., Lepage T., Blanquart S.I. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics.* 2009;25:2286-2288
12. Miller M.A., Pfeiffer W., Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Gateway Computing Environments Workshop (GCE) IEEE. 2010;1-8
13. Tekle Y.I., Wang F., Wood F.C., Anderson O.R., Smirnov A. New insights on the evolutionary relationships between the major lineages of Amoebozoa. *Sci Rep.* 2022;12:11173. doi 10.1038/s41598-022-15372-7
14. Kamyshatskaya O.G., Bondarenko N.I., Mesentsev Y.S., Chistyakova L.V., Nassonova E.S., Smirnov A.V. Molecular phylogeny of Polychaos annulatum (Amoebozoa, Tubulinea, Euamoebida) shows that genus Polychaos belongs to the family Hartmannellidae. *J Eukaryot Microbiol*. 2020;67:321-326. doi 10.1111/jeu.12782

6 Симпозиум «Генетика/ геномика, биоинформатика и системная биология растений»

Symposium “Genetics/ genomics, bioinformatics and systems biology of plants”

doi 10.18699/bgrs2024-6-1

Применение технологий параллельного и нанопорового секвенирования для изучения структурной организации митохондриального генома сои (*Glycine max* L. Merr.)

Александрович В.В.*, Синявская М.Г.
Институт генетики и цитологии НАН Беларуси, Минск, Беларусь
** valeria.alexandrovich@gmail.com*

Ключевые слова: митохондриальный геном; соя; полные нуклеотидные последовательности; секвенирование нового поколения; нанопоровое секвенирование

Мотивация и цель: Концепция существования митохондриального генома высших растений в виде «master circle» или мастер-хромосомы ставится под сомнение в последние годы [1]. Все больше доказательств появляется в пользу динамической структуры растительной митохондриальной ДНК (мтДНК), которая представлена в клетке множеством линейных и кольцевых молекул, возникающим в процессе рекомбинации и перестроек генома [2].

Митохондриальный геном сои (*Glycine max* L. Merr) считается одним из сложнейших геномов среди культурных растений. В геноме *G. max* присутствует необычайно высокая доля длинных повторов, рекомбинация по которым приводит к возникновению, предположительно, более тысячи разных сублимонов [3].

В базах данных нуклеотидных последовательной имеется три сборки мтДНК культурной сои: сортов Aiganhuang [4] и Zhonghuang 13 [5], митохондриальный геном которых, согласно авторам, представляет собой единственную кольцевую молекулу, и мтДНК сорта Williams 82 [3], существующей, предположительно, в виде двух кольцевых хромосом. Однако данные последовательности мтДНК не отражают возможного состояния генома митохондрий в живой клетке.

В данном исследовании была поставлена цель изучить полную нуклеотидную последовательность митохондриального генома культурной сои (*G. max*) с использованием двух методов секвенирования с учетом ее возможной лабильной структуры *in vivo*.

Методы и алгоритмы: Для сборки митохондриального генома *G. max* использовалось сочетание параллельного (Illumina MiSeq) и нанопорового (MinION, Oxford Nanopore Nechnologies) секвенирования. Полученные после параллельного секвенирования FASTQ-файлы были обработаны Trimmomatic, далее были собраны контиги с использованием SPAdes, контиги визуализировались с помощью Bandage. Combinator-FQ использовался для предварительной оценки копийности контигов, основываясь на их покрытии, и порядка их расположения.

Контиги из длинных прочтений нанопорового секвенирования были получены с помощью flye и canu. Выравнивание контигов паралелльного секвенирования на контиги длинных прочтений (SnapGene) было использовано для оценки возможного расположения блоков мтДНК относительно друг друга и идентификации повторов в последовательности митохондриального генома.

Результаты: В результате обработки данных NGS было получено 23 контига размером от 196 до 65372 п.н., различного покрытия и GC-состава (от 36 до 46 %). Один из контигов образует кольцевую молекулу размером 62661 п.н., не имеющую областей гомологии с остальными последовательностями, что говорит о существовании митохондриального генома сои в виде как минимум двух хромосом, не рекомбинирующих друг с другом. Замкнутость данного контига подтверждается также наличием длинных нанопоровых прочтений на границе соединения его в кольцо.

Выравнивание остальных 22 контигов на нанопоровые прочтения показало их существование в виде "блоков" мтДНК, расположенных в различной последовательности и ориентации относительно друг друга.

Контиги, имеющее высокое по сравнению с остальными покрытие, представлены в геноме митохондрий в виде двух и более копий. Предположительно, данные многокопийные участки размером от 440 до 12042 п.н. образуют рекомбинационные повторы, которые и обеспечивают множество обнаруженных конформаций мтДНК.

В мтДНК сои обнаружены инверсии размером до 183678 п.н., делеции и вставки участков генома. Один из контигов размером в 15259 п.н. отличается наличием двух открытых рамок считывания, кодирующих ДНК-полимеразу группы B и РНК-полимеразу, что имеет некоторое сходство с обнаруженными ранее в митохондриях растений линейными плазмидами [6]. В случае сои данная плазмида не имеет длинных инвертированных повторов и встроена в мтДНК. Два контига (196 п.н. и 1318 п.н.), по предварительным данным, образуют блок мтДНК, существующий во встроенном в хромосому виде и одновременно с этим находится в митохондриях в виде автономной кольцевой молекулы. Природа данного участка на данный момент находится под вопросом.

Всего, если представить одну из хромосом мтДНК в виде «master circle», митохондриальный геном сои имеет общий размер в 434295 п.н., GC-состав 44,94% и организован в виде двух хромосом, что соотносится с данными Liu и соавторов (сборка мтДНК сорта Williams 82) [3].

Выводы: Митохондриальный геном сои представлен *in vivo* минимум двумя молекулами, одна из которых имеет консервативную структуру, а вторая подвержена постоянным рекомбинационным перестройкам благодаря наличию длинных повторов в последовательности мтДНК и существует в клетке в виде разнообразных субгеномных молекул. Сочетание технологий Illumina и Oxford Nanopore позволило с большей уверенностью говорить о структуре мтДНК *G. max* и показало себя как перспективный метод сборки сложных митохондриальных геномов растений.

На данный момент стоит задача создания и опубликования в базе данных NCBI GenBank нуклеотидной последовательности митохондриального генома *G. max*, отражающей копийность блоков мтДНК и особенности ее структуры. Также в дальнейших исследованиях планируется изучить выявленные плазмидо-подобные молекулы.

Финансирование: Исследование поддержано ГПНИ «Биотехнологии 2» 2021–2025 гг. подпрограммы «Геномика, эпигеномика, биоинформатика» (рег. № НИОКТР 20210443).

Application of parallel and nanopore sequencing technologies for the study of soybeans (*Glycine max* L. Merr.) mitochondrial genome structural organization

Aleksandrovich V.V.*, Siniauskaya M.G.
Institute of Genetics and Cytology, NAS, Minsk, Belarus
** valeria.alexandrovich@gmail.com*

Key words: mitochondrial genome; soybean; complete nucleotide sequences; next-generation sequencing; nanopore sequencing

Motivation and Aim: The master circle model of higher plant mitochondrial genome has been questioned in recent years [1]. More and more evidence is emerging in support of dynamic structure of plant mitochondrial DNA (mtDNA), which in the living cell takes the form of a variety of linear and circular molecules that arise during the process of recombination and genome rearrangements [2].
The mitochondrial genome of soybean (*Glycine max* L. Merr) is considered to be one of the most complex genomes among cultivated plants. The *G. max* mtDNA contains an unusually high number of long repeats, the recombination of which leads to the formation of more than a thousand different subgenome molecules [3].
Nucleotide sequence databases contain three assemblies of mtDNA of cultivated soybean. According to the authors, the mitochondrial genome of cultivars Aiganhuang [4] and Zhonghuang 13 [5] is the single circular molecule, and mtDNA of cultivar Williams 82 [3] exists in the form of two circular chromosomes. However, these assemblies do not reflect the possible state of the mitochondrial genome in the living cell.
The goal of this study was to investigate the complete nucleotide sequence of cultivated soybeans mitochondrial genome using two sequencing methods. Also the possible labile structure of mtDNA *in vivo* was taken in consideration.
Methods and Algorithms: A combination of parallel (Illumina MiSeq) and nanopore (MinION, Oxford Nanopore Technologies) sequencing methods was used to assemble the G. max mitochondrial genome. The FASTQ files obtained after next-generation sequencing were processed by Trimmomatic, after that contigs were assembled using SPAdes. Contigs were visualized using Bandage. Combinator-FQ script was used to preevaluate the order and copy number of contigs based on their coverage.
Contigs from nanopore sequencing long reads were obtained using flye and canu. Alignment of parallel sequencing contigs to long-read contigs (SnapGene) was used to estimate the possible order of mtDNA blocks and identify repeats in the mitochondrial genome sequence.
Results: As a result of NGS data processing 23 contigs ranging in size (from 196 to 65372 bp) and GC-content (36 to 46%) were obtained. One of the contigs forms a circular molecule of 62661 bp in size, which has no homologous regions with other sequences, suggesting the existence of the soybean mitochondrial genome in the form of at least two chromosomes that do not recombinate with each other. The presence of long nanopore reads that contain the junction of two ends of this contig also confirms the formation of one circular molecule.

Alignment of the remaining 22 contigs to nanopore reads showed their existence in the form of blocks of mtDNA located in different order and orientation. Contigs with higher coverage are represented in the mitochondrial genome by two or more copies. Presumably, these multicopy regions ranging in size from 440 to 12042 bp form recombination repeats, which provoke the formation of multiple mtDNA conformation variants that were detected.

Inversions up to 183678 bp in size, deletions and insertions of genomic regions were observed in soybean mtDNA. One of the contigs 15259 bp long is notable for the presence of two open reading frames encoding group B DNA polymerase and RNA polymerase, which is of some similarity with linear plasmids previously described in plant mitochondria [6]. In the case of soybean, this plasmid does not have long inverted repeats and is integrated into mtDNA. Two contigs (196 bp and 1318 bp), according to preliminary data, form a mtDNA block that was found to be integrated into the chromosome and, at the same time, to be located in mitochondria as an autonomous circular molecule. The nature of this mtDNA block is currently in question.

Overall, the soybean mitochondrial genome has a total size of 434295 bp, a GC-content of 44,94% and is organized in the form of two chromosomes, which correlates with the data of Liu et al. (assembly of mtDNA of Williams 82 cultivar) [3].

Conclusion: The soybean mitochondrial genome is represented *in vivo* by at least two molecules, one of which has a conservative structure, and the second is subjected to constant recombinational rearrangements due to the presence of long repeats in the mtDNA sequence. This chromosome exists in the cell in the form of various subgenomic molecules. The combination of Illumina and Oxford Nanopore technologies allowed us to properly investigate the structure of *G. max* mtDNA and showed itself to be a promising method for assembling complex plant mitochondrial genomes.

At present, our goal is to obtain and publish in the NCBI GenBank database the full nucleotide sequence of the cultivated soybeans mitochondrial genome, which reflects the copy number and structural peculiarities of mtDNA blocks. We are also planning to investigate the nature of detected plasmid-like molecules in further studies.

Funding: The study is supported by the State Scientific Research Program «Biotechnologies 2» 2021–2025, Subprogram «Genomics, epigenomics, bioinformatics» (No. 20210443).

Список литературы/References

1. Михайлова Ю.В., Терентьева Л.Ю. Гигантские митохондриальные геномы высших растений. *Успехи современной биологии*. 2017;137(3):237–246.
[Mikhaylova Yu.V., Terent'eva L.Yu. Huge Mitochondrial Genomes in Embryophyta Plants. *Uspehi sovremennoy biologii = Advances in modern biology*. 2017;137(3):237–246 (in Russian)]
2. Kozik A. et al. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLoS Genet*. 2019;15(8):e1008373.
3. Liu H. et al. Structural variation of mitochondrial genomes sheds light on evolutionary history of soybeans. *The Plant Journal*. 2021;108:1456–1472.
4. Chang S. et al. The Mitochondrial Genome of Soybean Reveals Complex Genome Structures and Gene Evolution at Intercellular and Phylogenetic Levels. PLoS ONE. 2013;8(2):e56502.
5. Shen Y. et al. Update soybean Zhonghuang 13 genome to a golden reference. *Sci China Life Sci*. 2019;62(9):1257-1260.
6. Handa H., Itani K., Sato H. Structural features and expression analysis of a linear mitochondrial plasmid in rapeseed (*Brassica napus* L.). *Mol Genet Genomics*. 2002;267:797–805.

doi 10.18699/bgrs2024-06-01

Морфометрия колосьев пшеницы на основе анализа цифровых изображений и методов глубокого машинного обучения

Афонников Д.А.[1, 2, 3]*, Генаев М.А.[1, 2], Комышев Е.Г.[1], Артеменко Н.В.[3], Бусов И.Д.[3], Епифанов Р.Ю.[3], Кручинина Ю.В.[1, 2], Коваль В.С.[1, 2], Пискарев В.В.[4], Гончаров Н.П.[1]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

[3] *Новосибирский государственный университет, Новосибирск, Россия*

[4] *СибНИИРС – филиал Института цитологии и генетики СО РАН, Краснообск, Новосибирская область, Россия*

* *ada@bionet.nsc.ru*

Ключевые слова: пшеница; колос, фенотипирование; анализ цифровых изображений; глубокое машинное обучение

Мотивация и цели: Морфометрические характеристики колоса пшеницы являются одними из наиболее важных для генетиков и селекционеров, поскольку тесно связаны с такими хозяйственно ценными качествами, как продуктивность, отсутствие ломкости колоса и легкость обмолота. Для выявления генов, контролирующих данные признаки, используются статистические методы (QTL, GWAS), для успешности применения которых необходимы сбор и анализ большого количества фенотипических данных. В настоящее время в большинстве исследований фенотипирование выполняется экспертами на основании визуального анализа колоса и измерений вручную, что является трудозатратным. Для повышения эффективности фенотипирования колосьев необходимы методы, основанные на автоматическом анализе изображений.

Методы и алгоритмы: Для фенотипирования колосьев пшеницы в нашем коллективе были разработаны методы на основе анализа двумерных изображений. Эти методы используют как методы компьютерного зрения, так и глубокого машинного обучения и позволяют выделять на изображении области колоса и остей, выделять контур тела колоса, определять его геометрические параметры, подсчитывать число колосков в колосе, оценивать цвет колоса и остей, производить классификацию колосьев по опушению чешуй и плоидности растения.

Результаты: Разработанные методы были использованы для оценки характеристик колосьев разных видов пшеницы из коллекции Н.П. Гончарова и коллекции мягкой пшеницы СибНИИРС.

Выводы: Использование методов автоматического анализа цифровых изображений колосьев позволяет проводить массовый анализ характеристик колосьев пшеницы с высокой производительностью и точностью.

Финансирование: Работа выполнена при поддержке РНФ, проект № 23-14-00150. Вычисления проводились с использованием ресурсов ЦКП «Биоинформатика».

Morphometry of wheat spikes based on digital image analysis and deep machine learning methods

Afonnikov D.A.[1, 2, 3]*, Genaev M.A.[1, 2], Komyshev E.G.[1], Artemenko N.V.[3], Busov I.D.[3], Epifanov R.Yu.[3], Kruchinina Yu.V.[1, 2], Koval V.S.[1, 2], Piskarev V.V.[4], Goncharov N.P.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

[4] *Siberian Research Institute of Plant Production and Breeding – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *ada@bionet.nsc.ru*

Key words: wheat; spike; phenotyping; digital image analysis; deep machine learning

Motivation and Aim: The morphometric characteristics of a wheat spike are among the most important for geneticists and breeders, since they are closely related to such economically valuable qualities as productivity, lack of fragility of the ear and ease of threshing. Statistical methods (QTL, GWAS) are used to identify the genes that control these traits, for the success of which it is necessary to collect and analyze a large amount of phenotypic data. Currently, in most studies, phenotyping is performed by experts based on visual analysis of the ear and manual measurements, which is labor-intensive. To increase the efficiency of ear phenotyping, methods based on automatic image analysis are needed.

Methods and Algorithms: For the phenotyping of wheat spike, we developed methods based on the analysis of two-dimensional images. These methods use both computer vision and deep machine learning methods and allow to determine the areas of the spike and awns in the image, determine the contour of the ear body, determine its geometric parameters, count the number of spikelets in the spike, evaluate the color of the spike and awns, classify spikes according to the glume pubescence and ploidy of the plant.

Results: The developed methods were used to evaluate the characteristics of ears of different types of wheat from the collection of N.P. Goncharov and the collection of bread wheat from SibRIPP&B.

Conclusion: The use of methods for automatic analysis of digital images of wheat spikes allows for a mass analysis of their characteristics with high productivity and accuracy.

Funding: The work was carried out with the support of the RSF, project No. 23-14-00150. The data analysis was carried out using the resources of the Joint Computational Center “Bioinformatics”.

doi 10.18699/bgrs2024-06-02

Разработка вычислительного конвейера для поиска семейств мультидоменных белков в геномах растений

Бочарникова М.[1, 2]*, Турнаев И.[1], Афонников Д.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

* *lachynova@bionet.nsc.ru*

Ключевые слова: вычислительный конвейер; мультидоменные белки; филогения

Мотивация и цель: Современные технологии секвенирования привели к огромному потоку данных о последовательностях генома растений. Для определения функций белка эти данные требуют высокопроизводительной компьютерной аннотации. Целью настоящей работы являлось создание вычислительного конвейера OrthoDom, который предназначен для поиска ортологичных белков с учетом их доменного состава.

Методы и алгоритмы: Конвейер реализован на основе контейнерной системы Snakemake [1]. В качестве входных данных конвейер принимает протеомы организмов, у которых пользователь хочет найти белки, наборы референсных белков, для которых известны функции и домены, а также профили HMM [2] для доменов этих белков. Конвейер выводит список ортологичных последовательностей для эталонных белков, а также информацию об их доменном составе и филогенетических деревьях для идентифицированных ортологичных семейств.

Результаты: Мы протестировали точность системы и внедрили ее для анализа растительных белков. Конвейер использовался для идентификации ортологов в геномах злаков.

Выводы: Разработанный нами вычислительный конвейер может успешно применяться для более эффективного поиска семейств мультидоменных белков в геномах растений.

Финансирование: Исследование поддержано Курчатовским геномным центром ИЦиГ СО РАН (проект № 075-15-2019-1662).

Development of a computational pipeline for searching for families of multidomain proteins in plant genomes

Bocharnikova M.[1, 2]*, Turnaev I.[1], Afonnikov D.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *lachynova@bionet.nsc.ru*

Key words: computational pipeline; multi-domain proteins; phylogeny

Motivation and Aim: Modern sequencing technologies have led to a huge flow of data on plant genome sequences. These data require high-performance computer annotation

to determine protein functions. The purpose of this work was to create the Orthodox computing pipeline, which is designed to search for orthologous proteins taking into account their domain composition.

Methods and Algorithms: The pipeline is implemented based on the Snakemake container system [1]. As input, the pipeline accepts proteomes of organisms in which the user wants to find proteins, sets of reference proteins for which functions and domains are known, as well as HMM profiles [2] for the domains of these proteins. The pipeline outputs a list of orthologous sequences for reference proteins, as well as information about their domain composition and phylogenetic trees for identified orthologous families.

Results: We tested the accuracy of the system and implemented it for the analysis of plant proteins. The conveyor was used to identify orthologs in the genomes of cereals.

Conclusion: The computational pipeline we have developed can be successfully used to more efficiently search for families of multidomain proteins in plant genomes.

Funding: The study was supported by the Kurchatov Genomic Center of the ICG SB RAS (project No. 075-15-2019-1662).

Список литературы/References

1. Köster J., Rahmann S. Snakemake – a scalable bioinformatics workflow engine. *Bioinformatics*. 2012;28(19):2520-2522
2. Eddy S. R. Hidden markov models. *Curr. Opin. Struct. Biol*. 1996;6(3):361-365

doi 10.18699/bgrs2024-06-03

Идентификация потенциальных промоторных последовательностей, предсказанных методом MADHS, путем анализа доступности хроматина и поиска неаннотированных сайтов старта транскрипции

Бубнова А.*, Гайдукова С., Яковлева И., Камионская А.
ФГУ «Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук», Москва, Россия
** an_bubnova@mail.ru*

Ключевые слова: предсказанный потенциальный промотор; сайт начала транскрипции; доступная область хроматина

Мотивация и цель: Поиск, предсказание и идентификация неаннотированных промоторных последовательностей может иметь фундаментальное значение для обнаружения в геномах функциональных областей, не аннотированных ранее. Но на данный момент существующие алгоритмы прогнозирования промоторов не исключают появления ложноположительных результатов, что является серьезной проблемой вычислительной биологии. В связи с этим при наличии предсказанных промоторных последовательностей целесообразным является проведение дополнительной проверки их свойств для подтверждения возможности их функционирования в геноме в роли регуляторных элементов. Целью данной работы была проверка предсказанных методом MADHS [1] промоторных последовательностей 1-й хромосомы генома риса *O. sativa* на наличие потенциальных неаннотированных сайтов старта транскрипции в районе 3’-конца предсказанных промоторов, а также изучение доступности хроматина в районе предсказанных последовательностей.
Методы и алгоритмы: Предсказанные методом MADHS промоторные последовательности были выгружены из Базы данных потенциальных промоторных последовательностей (доступно по адресу: http://victoria.biengi.ac.ru/cgi-bin/dbPPS/index.cgi), всего было отобрано 126 последовательностей из хромосомы 1 генома риса с наибольшим аргументом нормального распределения, который показывает степень неслучайности созданного выравнивания при предсказании. В геноме эти последовательности расположены на расстоянии не ближе 1000 п. н. от известных генов и не ближе, чем на 200 п. н. от транспозонов и SINE повторов. Это гарантирует, что предсказанные потенциальные промоторы не связаны с транспозонами и промоторами известных генов. Длина выбранных предсказанных потенциальных промоторов варьирует от 428 до 599 нуклеотидов, а расстояние от них до ближайших аннотированных транскриптов – от 1044 до 92 434 п. н.
Поиск транскрибируемых участков ДНК (потенциальных неаннотированных сайтов старта транскрипции) за потенциальными промоторными последовательностями проводился путем анализа результатов секвенирования кэп-анализов экспрессии генов (CAGE-seq) риса *O. sativa*, доступных в SRA NCBI. Данный анализ позволяет картировать большинство сайтов начала транскрипции

и, как следствие, их промоторов. Адаптеры и чтения низкого качества в необработанных данных секвенирования удалялись с помощью Trim-Galore. Выравнивание отфильтрованных данных на геном первой хромосомы риса проводилось с помощью *bowtie2* с опцией *very-sensitive.* Полученный в результате bam-файл был конвертирован в sam-файл, проиндексирован и отсортирован с помощью инструментов Samtools, затем из этого файла были получены пики, характеризующие участки потенциальных сайтов старта транскрипции. Расстояние от них до предсказанных промоторов оценивали с использованием Bedtools closest. В результате отбирались предсказанные промоторы, за которыми в направлении 3’ были обнаружены потенциальные сайты начала транскрипции в пределах 1000 п. н. Визуализация выбранных данных осуществлялась с помощью PyGenomeTracks.

Структура хроматина анализировалась путем изучения данных секвенирования ATAC-seq, также взятых из SRA NCBI, позволяющих распознать открытые участки хроматина, доступные для посадки факторов транскрипции. Такие участки могут помогать в идентификации промоторных последовательностей. Алгоритм действий с данными ATAC-seq был идентичен вышеописанной работе с данными CAGE-seq, за исключением последнего этапа: после выявления пиков, характеризующих открытую область хроматина, отбирались промоторы, имеющие пересечение с пиками, тогда как расстояние до ближайших пиков не изучалось.

Результаты: При анализе данных CAGE-seq было обнаружено наличие четких пиков, характерных для областей сайтов старта транскрипции, расположенных в свободных от аннотированных транскриптов участках генома. Это может свидетельствовать о возможном существовании ранее неаннотированных транскриптов в геноме риса. Достоверность полученных результатов была подтверждена наличием пиков CAGE-seq в районах 5’-концов аннотированных транскриптов. Из 126 отобранных ранее предсказанных промоторных последовательностей было выявлено 14 промоторов, имеющих за собой потенциальные сайты старта транскрипции:

- 3 потенциальных TSS расположены не далее 100 п.н. от 3'-концов предсказанных промоторов;
- 9 потенциальных TSS, пересекающихся с 3'-концами предсказанных промоторов;
- 2 потенциальных TSS, расположенных на расстоянии от 100 до 1000 п. н. от 3’-концов предсказанных промоторов.

В результате анализа ATAC-seq из 126 предсказанных промоторов было отобрано 16 последовательностей, пересекающихся с пиками ATAC-seq. Это указывает на то, что предсказанные промоторы расположены в открытых участках хроматина, доступных для посадки транскрипционных факторов.

Выводы: Из 126 первоначально отобранных предсказанных промоторных последовательностей: 14 предсказанных промоторов имеют за собой потенциальную транскрипцию и 16 находятся в открытой области хроматина. Только 4 предсказанных промотора удовлетворяют обоим условиям проверки. Данные последовательности наиболее вероятно являются промоторными областями и рекомендованы к дальнейшей экспериментальной проверке *in vivo*.

Финансирование: Исследование выполнено при поддержке Минобрнауки России в рамках соглашения № 075-15-2022-318 от 20.04.2022 о предоставлении гранта в

форме субсидий из федерального бюджета на осуществление государственной поддержки создания и развития научного центра мирового уровня «Агротехнологии будущего».

Identification of potential promoter sequences predicted by the MADHS method by analyzing chromatin accessibility and searching for unannotated transcription start sites

Bubnova A.*, Gaidukova S., Yakovleva I., Kamionskaya A.

Federal State Institution "Federal Research Centre "Fundamentals of Biotechnology" of the Russian Academy of Sciences", Moscow, Russia

** an_bubnova@mail.ru*

Key words: potential predicted promotor; transcription start site; chromatin accessibility

Motivation and Aim: Search, prediction, and identification of unannotated promoter sequences can play a fundamental role in the discovery of functional regions in genomes that have not been previously annotated. However, currently existing promoter prediction algorithms do not exclude the occurrence of false positives, which is a serious problem in computational biology. Therefore, in the presence of predicted promoter sequences, it is reasonable to carry out additional testing of their properties to confirm the possibility of their functioning in the genome as regulatory elements. The aim of this work was to check the predicted by the MADHS method [1] promoter sequences of the 1st chromosome of the rice *O. sativa* genome for the presence of potential unannotated transcription start sites near their 3'-end, as well as to study chromatin accessibility in the region of the predicted sequences.

Methods and Algorithms: Promoter sequences predicted by the MADHS method were downloaded from the Potential Promoter Sequences Database (available at http://victoria.biengi.ac.ru/cgi-bin/dbPPS/index.cgi), and 126 sequences with the largest normal distribution argument were selected from chromosome 1 of the rice genome, indicating the degree of nonrandom alignment created during the prediction process. In the genome, these sequences are located no closer than 1,000 bp from known genes and no closer than 200 bp from transposons and SINE repeats. This ensures that the predicted potential promoters are not linked to transposons and promoters of known genes. The length of the selected predicted potential promoters ranges from 428 to 599 nucleotides, and the distance from them to the nearest annotated transcripts ranges from 1,044 to 92,434 bp.

A search for transcribed DNA regions (potential unannotated transcription start sites) behind potential promoter sequences was performed by analyzing the results of gene expression sequencing analysis (CAGE-seq) of rice *O. sativa* available at NCBI SRA. This analysis allows mapping of most transcription start sites and, as a consequence, their promoters. Adapters and low-quality reads in the raw sequencing data were removed using the Trim-Galore program. Alignment of the filtered data to the genome of the first rice chromosome was performed using the bowtie2 program with the -very-sensitive option. The resulting bam file was converted to a sam file, indexed and sorted using Samtools tools, then peaks characterizing regions of potential transcription start sites were extracted. The distance from these to predicted promoters was estimated using

Bedtools closest. As a result, predicted promoters were selected, behind which potential transcription start sites within 1000 bp were detected in the 3' direction. The selected data were visualized using PyGenomeTracks.

Chromatin structure was analyzed using ATAC-seq sequencing data, also taken from the NCBI SRA, which allows us to recognize open regions of chromatin available for landing transcription factors. Such regions can help in the identification of promoter sequences. The algorithm of actions with ATAC-seq data was identical to the work with CAGE-seq data described above, except for the last step: after identifying the peaks characterizing the open region of chromatin, promoters overlapping with the peaks were selected, and the distance to the nearest peaks was not studied.

Results: Analysis of CAGE-seq data revealed the presence of distinct peaks characteristic of transcription start site regions located in regions of the genome that do not contain annotated transcripts. This may indicate the possible existence of previously unannotated transcripts in the rice genome. The validity of the results was confirmed by the presence of CAGE-seq peaks in the 5'-terminal regions of annotated transcripts. Out of 126 selected previously predicted promoter sequences, 14 promoters with potential transcription start sites behind them were identified:

- 3 potential TSS located no further than 100 bp from the 3'-end of the predicted promoters;
- 9 potential TSSs overlapping with the 3'-ends of predicted pro-motors;
- 2 potential TSS located between 100 and 1000 bp from the 3'-end of predicted promoters.

ATAC-seq analysis selected 16 sequences that overlapped with ATAC-seq peaks from 126 predicted promoters. This indicates that the predicted promoters are located in open regions of chromatin accessible for seeding of transcription factors.

Conclusion: Of the 126 previously predicted promoter sequences selected: 14 predicted promoters have potential transcription behind them and 16 are located in the open region of chromatin. Only 4 predicted promoters satisfy both validation conditions. These sequences are the most likely promoter regions and are recommended for further experimental verification *in vivo*.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement No. 075-15-2022-318 date April 20, 2022 on providing a grant in the form of subsidies from the Federal budget of Russian Federation. The grant was provided for state support for the creation and development of a World-class Scientific Center “Agrotechnologies for the Future”.

Список литературы/References

1. Korotkov E.V. et al. Mathematical Algorithm for Identification of Eukaryotic Promoter Sequences. *Symmetry*. 2021;13(6):917

doi 10.18699/bgrs2024-06-04

Локусы лизин-специфических деметилаз у растений способны транскрибировать кольцевые молекулы РНК

Бурсаков С.А.*, Драгни А.Г., Ефремов Л.Н., Князев А.Н., Дивашук М.Г.
ФГБНУ «Всероссийский научно-исследовательский институт сельскохозяйственной биотехнологии (ФГБНУ ВНИИСБ), Москва, Россия
** sergeymoscu@gmail.com*

Ключевые слова: деметилирование гистона H3K27me3; кольцевые РНК; эпигенетика; *Arabidopsis thaliana*; *Glycine max*; *Solanum lycopersicum*; *Triticum aestivum*; *Zea mays*

Мотивация и цель: Метилирование гистонов является одним из ключевых эпигенетических механизмов, влияющих на рост и развитие растений в различных условиях. Этот процесс включает добавление метильных групп к аминокислотным остаткам в гистонах, что изменяет структуру хроматина и регулирует доступность ДНК для ферментов, участвующих в транскрипции. Существует множество видов метилирования гистонов, среди которых метилирование лизина 27 в гистоне H3 (H3K27me3) представляет собой яркий пример проявления сайленсинга генов, являющийся консервативным в царстве животных и растений [1]. Деметилирование H3K27me3 – обратный процесс, удаление метильных групп с лизина, что приводит к релаксации хроматина и увеличению транскрипции генов. Лизин-специфические деметилазы – белки, осуществляющие деметилирование. В настоящее время известно участие генов деметилаз в процессах роста и развития (прорастание семян, цветение), формировании стрессовых ответов на абиотические факторы, такие как засуха и засоление [2]. У арабидопсиса показано участие генов деметилаз в переходе к цветению, в формировании солеустойчивости, в ответе к повышенной температуре, в прорастании семян на свету и др. [3–6]. У томата показано участие Histone H3K27me3 demethylase SlJMJ3 в созревании плодов [7], а у сахарного тростника – как ответ на засуху [8]. Таким образом, деметилирование гистонов представляет собой мощный инструмент для улучшения агрономических характеристик и устойчивости сельскохозяйственных культур.

Кольцевые РНК (circRNA) участвуют, вероятно, в механизме регуляции экспрессии генов деметилаз. CircRNA – это класс РНК, образованных из прекурсорных молекул мРНК. Их общим свойством является образование кольцевой структуры. Изучение кольцевых РНК у растений привлекает внимание исследователей прежде всего из-за их возможности регулировать экспрессию генов. Возможным механизмом их действия является способность некоторых кольцевых РНК выступать губками миРНК, тем самым освобождая потенциальные мишени протеинкодирующих генов миРНК. Кроме того, показано, что кольцевые РНК могут регулировать экспрессию родительского гена, связываясь с промоторным регионом. Немаловажным является свойство кольцевых РНК влиять на трансляцию линейных мРНК, повышая или снижая синтез целевого белка. И наконец, кольцевые РНК могут быть мишенями для РНК-связывающих белков, выступая конкурентами для их линейных аналогов. Разнообразие механизмов действия кольцевых РНК позволяет влиять на

различные биологические процессы в растениях. Ранее была показана роль некоторых РНК при росте, развитии и формировании ответов на стрессовые факторы у арабидописа, томата, риса и др. [9, 10]. Однако влияние кольцевых РНК на механизм метилирования гистонов не выявлено.

Целью работы был поиск и анализ кольцевых РНК, расположенных в локусах генов гистоновых деметилаз у растений *Arabidopsis thaliana*, *Glycine max*, *Solanum lycopersicum*, *Triticum aestivum* и *Zea mays.*

Методы и алгоритмы: В работе мы использовали базу данных PlantcircBase 7.0, содержащую информацию по предсказанным circRNA, с их нуклеотидными последовательностями, локализацией на геномах и возможном взаимодействии с miRNA для растений арабидобсиса, томата и мягкой пшеницы [11]. Аннотированные геномы были взяты из Phytozome [12]. Пересечение локусов кольцевых РНК с аннотированными участками геномов использовали инструменты bedtools и samtools. Визуализация и обработка данных проводились в R studio соответствующими пакетами работы с табличными данными. Для работы с нуклеотидными последовательностями извлеченных из баз данных и построения филогенетических деревьев использовали Ugene [13]. Для валидации кольцевых РНК использовали ОТ-ПЦР с подбором праймеров на линейные и кольцевые молекулы с использованием Primer3. Выделение РНК, обратную транскрипцию и ПЦР проводили по стандартной методике, описанной ранее [14]. Растения выращивали в климатических камерах 16/8 светового дня и температурой 25 °C.

Результаты: Анализ существующих геномных баз данных и имеющихся предсказанных кольцевых РНК растений арабидопсиса, томата, пшеницы, сои и кукурузы показал, что в генах лизин-специфических деметилаз могут быть сайты обратного сплайсинга, которые указывают на возможность образования кольцевых РНК. Сайты обратного сплайсинга имеются не во всех локусах гистоновых деметилаз у рассматриваемых растений. Доказательства формирования кольцевых структур с использованием ОТ-ПЦР получены для некоторых кольцевых РНК выбранных нами растений. Показана низкая экспрессия кольцевых РНК локусов лизин-специфических деметилаз по сравнению с линейным аналогом. Такая экспрессия может быть объяснена отсутствием стрессовых условий выращивания растений. Некоторые кольцевые РНК генов гистоновых деметилаз могут выступать мишенями миРНК. Выявлена низкая консервативность среди проанализированных кольцевых РНК, расположенных в генах лизин-специфических деметилаз.

Выводы: Отдельные гены лизин-специфических деметилаз у арабидопсиса, томата, сои, пшеницы и кукурузы могут транскрибировать кольцевые молекулы РНК. Наличие таких молекул указывает на возможность регуляции экспрессии соответствующих генов кольцевыми структурами, приводящей к эпигенетическим изменениям. Планируемые исследования влияния кольцевых молекул РНК на уровень деметилирования H3K27me позволят разработать новый механизм регулирования роста, развития и повышения устойчивости сельскохозяйственных растений к стрессовым факторам.

Финансирование: Исследование поддержано государственным заданием FGUM-2022-0007.

Lysine-specific demethylase loci in plants are capable of transcribing circRNA molecules

Bursakov S.A.*, Dragni A.G., Efremov L.N., Knyazev A.N., Divashuk M.G.
Federal State Budgetary Scientific Institution "All-Russia Research Institute of Agricultural Biotechnology", Moscow, Russia
** sergeymoscu@gmail.com*

Key words: histone H3K27me3 demethylation; circular RNAs; epigenetics; *Arabidopsis thaliana*; *Glycine max*; *Solanum lycopersicum*; *Triticum aestivum*; *Zea mays*

Motivation and Aim: Histone methylation is one of the key epigenetic mecha- nisms affecting plant growth and development under various conditions. This process involves the addition of methyl groups to amino acid residues in histones, which changes the structure of chromatin and regulates the availability of DNA for enzymes involved in transcription. There are many types of histone methylation, among which the methylation of lysine 27 in histone H3 (H3K27me3) is a striking example of gene silencing manifestation that is conserved in the animal and plant kingdom [1]. Demethylation of H3K27me3 is the reverse process, removal of methyl groups from lysine, which leads to chromatin relaxation and increased gene transcription. Lysine-specific demethylases – proteins that carry out demethylation. It is now known that demethylase genes participate in the processes of growth and development (seed germination, flowering), formation of stress responses to abiotic factors such as drought and salinity [2]. In Arabidopsis, demethylase genes have been shown to be involved in the transition to flowering, in the formation of salt tolerance, in the response to elevated temperature, in seed germination under light, etc. [3–6]. In tomato, Histone H3K27me3 demethylase SlJMJ3 has been shown to be involved in fruit ripening [7], and in sugarcane as a response to drought [8]. Thus, histone demethylation is a powerful tool for improving agronomic traits and crop stability.

Circular RNAs (circRNAs) are probably involved in the mechanism of regulation of demethylase gene expression. CircRNAs are a class of RNAs formed from precursor mRNA molecules. Their common property is the formation of a ring structure. The study of circular RNAs in plants attracts the attention of researchers primarily because of their ability to regulate gene expression. A possible mechanism of their action is the ability of some circular RNAs to act as miRNA sponges, thereby releasing potential targets of protein-coding miRNA genes. In addition, it has been shown that circular RNAs can regulate the expression of the parental gene by binding to the promoter region. The property of circular RNAs to influence the translation of linear mRNAs by increasing or decreasing the synthesis of the target protein is also important. Finally, circular RNAs can act as targets for RNA-binding proteins, acting as competitors for their linear analogs. The diversity of mechanisms of action of circular RNAs makes it possible to influence various biological processes in plants. The role of some RNAs in the growth, development, and formation of responses to stress factors in Arabidopsis, tomato, rice, etc. was shown earlier [9, 10]. However, the influence of circular RNAs on the mechanism of histone methylation has not been revealed.

The aim of this work was to search for and analyze circular RNAs located in the loci of histone demethylase genes in Arabidopsis thaliana, Glycine max, Solanum lycopersicum, Triticum aestivum, and Zea mays plants.

Methods and Algorithms: In this work, we used PlantcircBase 7.0 database containing information on predicted circRNAs, with their nucleotide sequences, localization on genomes and possible interaction with miRNAs for Arabidopsis, tomato and soft wheat plants [11]. Annotated genomes were taken from Phytozome [12]. Intersection of circular RNA loci with annotated regions of the genomes was performed using bedtools and samtools tools. Data visualization and processing were performed in R studio using appropriate packages for working with tabular data. Ugene [13] was used to work with nucleotide sequences extracted from databases and to construct phylogenetic trees. To validate the circular RNAs, we used RT-PCR with primer selection on linear and circular molecules using Primer3. RNA isolation, reverse transcription, and PCR were performed according to the standard method described previously [14]. Plants were grown in climatic chambers of 16/8 daylight hours and 25 °C temperature.

Results: Analysis of existing genomic databases and available predicted circular RNAs from plants of Arabidopsis, tomato, wheat, soybean, and maize showed that there may be reverse splicing sites in lysine-specific demethylase genes that indicate the possibility of circular RNA formation. Reverse splicing sites are not present in all loci of histone demethylases in the plants under consideration. Evidence for the formation of ring structures using OT-PCR was obtained for some ring RNAs of our selected plants. Low expression of the ring RNA loci of lysine-specific demethylases was shown compared to the linine analog. This expression can be explained by the absence of stressful conditions of plant cultivation. Some circular RNAs of histone demethylase genes may act as miRNA targets. Low conservation among the analyzed circular RNAs located in the genes of lysine-specific demethylases was revealed.

Conclusion: Some genes of lysine-specific demethylases in arabidopsis, tomato, soybean, wheat, and maize can transcribe circular RNA molecules. The presence of such molecules indicates the possibility of regulation of expression of co-relevant genes by ring structures, leading to epigenetic changes. The planned studies of the influence of RNA ring molecules on the level of demethylation of H3K27me will make it possible to develop a new mechanism to regulate growth, development and increase the resistance of agricultural plants to stress factors.

Funding: The study is supported by the State Assignment FGUM-2022-0007.

Список литературы/References

1. Xiao J., Wagner D. Polycomb repression in the regulation of growth and development in Arabidopsis. *Curr Opin Plant Biol*. 2015;23:15-24. doi 10.1016/j.pbi.2014.10.003
2. Crevillén P. Histone demethylases as counterbalance to H3K27me3 silencing in plants. *iScience*. 2020;23(11):101715. doi 10.1016/j.isci.2020.101715
3. He K., Mei H., Zhu J., Qiu Q., Cao X., Deng X. The histone H3K27 demethylase REF6/JMJ12 promotes thermomorphogenesis in Arabidopsis. *Natl Sci Rev*. 2022;9(5):nwab213. doi 10.1093/nsr/nwab213
4. Yang W., Jiang D., Jiang J., He Y. A plant-specific histone H3 lysine 4 demethylase represses the floral transition in Arabidopsis. *Plant J*. 2010;62(4):663-673. doi 10.1111/j.1365-313X.2010.04182.x
5. Shen Y., Conde e Silva N., Audonnet L., Servet C., Wei W., Zhou D.-X. Over-expression of histone H3K4 demethylase gene JMJ15 enhances salt tolerance in Arabidopsis. *Front Plant Sci*. 2014;5:290. doi 10.3389/fpls.2014.00290
6. Wang Y., Gu D., Deng L., He C., Zheng F., Liu X. The histone H3K27 demethylase REF6 is a positive regulator of light-initiated seed germination in arabidopsis. *Cells*. 2023;12(2):295. doi 10.3390/cells12020295

7. Li Z. et al. Histone H3K27 demethylase SlJMJ3 modulates fruit ripening in tomato. *Plant Physiol*. 2024;kiae233. doi 10.1093/plphys/kiae233
8. Yu G. et al. H3K27 demethylase SsJMJ4 negatively regulates drought-stress responses in sugarcane. *J Exp Bot*. 2024;75(10):3040-3053. doi 10.1093/jxb/erae037
9. Liu R., Ma Y., Guo T., Li G. Identification, biogenesis, function, and mechanism of action of circular RNAs in plants. *Plant Commun*. 2023;4(1):100430. doi 10.1016/j.xplc.2022.100430
10. Chu Q. et al. Recent origination of circular RNAs in plants. *New Phytol*. 2021;233(1):515-525. doi 10.1111/nph.17798
11. Xu X. et al. PlantcircBase 7.0: Full-length transcripts and conservation of plant circRNAs. *Plant Commun*. 2022;3(4):100343. doi 10.1016/j.xplc.2022.100343
12. Goodstein D.M. et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(D1):D1178-D1186. doi 10.1093/nar/gkr944
13. Okonechnikov K., Golosova O., Fursov M., UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28(8):1166-1167. doi 10.1093/bioinformatics/bts091
14. Fesenko I. et al. Distinct types of short open reading frames are translated in plant cells. *Genome Res*. 2019;29(9):1464-1477. doi 10.1101/gr.253302.119

doi 10.18699/bgrs2024-06-05

Методы на основе анализа изображений для мониторинга грибных заболеваний и сорных растений в полевых условиях

Генаев М.А.[1, 3], Кожекин М.В.[1], Коваль В.С.[1, 3], Афонников Д.А.[1, 2, 3]*

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

* *ada@bionet.nsc.ru*

Ключевые слова: глубокое обучение; компьютерное зрение; сегментация; детекция

Мотивация и цель: Одна из актуальных проблем в крупных агрохолдингах – мониторинг посевов для выявления грибных заболеваний и очагов сорной растительности. Заболевания зерновых, вызванные патогенными грибами, могут существенно снизить урожайность сельскохозяйственных культур. Им подвержены многие культуры. Заболевание трудно контролировать в больших масштабах; мониторинг посевов помогает выявить заболевания и очаги сорной растительности на ранней стадии и принять меры по предотвращению их распространения. Цель работы – разработка метода на основе анализа изображений для мониторинга грибных заболеваний и сорных растений в полевых условиях.

Методы и алгоритмы: Одним из эффективных методов мониторинга является идентификация заболеваний на основе анализа цифровых изображений с возможностью получения их в полевых условиях, с использованием мобильных устройств. В работе предложен метод распознавания пяти грибных болезней побегов у злаков (листовая ржавчина, стеблевая ржавчина, желтая ржавчина, мучнистая роса и септориоз) как отдельно, так и при множественном заболевании, с возможностью выявления стадии развития растения. Был сформирован набор из 2414 изображений грибковых заболеваний пшеницы, для которых была проведена экспертная маркировка по типу заболевания. Более 80 % изображений в наборе данных соответствуют меткам одного заболевания, 12 % представлены здоровыми растениями, а 6 % изображений представлены несколькими заболеваниями. В процессе создания данного набора был применен метод уменьшения вырожденности обучающих данных на основе алгоритма хеширования изображений.

Для определения очагов сорной растительности в полевых условиях использовались БПЛА для получения геопространственных растровых файлов. Методы компьютерного зрения для реконструкции карты плотности посевов и определения контуров культурных и сорных растений.

Результаты: Был сформирован набор из 2414 изображений грибковых заболеваний пшеницы, для которых была проведена экспертная маркировка по типу заболевания. Более 80 % изображений в наборе данных соответствуют меткам одного заболевания, 12 % представлены здоровыми растениями, а 6 % изображений представлены несколькими заболеваниями. В процессе создания

данного набора был применен метод уменьшения вырожденности обучающих данных на основе алгоритма хеширования изображений. Алгоритм распознавания заболеваний основан на CNN с архитектурой EfficientNet. Наилучшую точность (0.942) показала сеть со стратегией обучения, основанной на дополнении и передаче стилей изображений. Метод распознавания реализован в виде бота на платформе Telegram, который позволяет пользователям оценивать растения по изображению в полевых условиях. Набор данных и модель выложены в открытый доступ на сайте http://wfd.sysbio.ru.

Для мониторинга полей и применения БПЛА нами разработана система SeedlingsNet для автоматической оценки количества всходов на RGB изображениях полей в формате GeoTIFF разрешением 1–2 см/пикс. Идентификация всходов широкорядных полевых культур осуществляется с точностью более 95 % по сравнению с ручной идентификацией всходов. Компания «ГеосАэро» планирует использовать полученный инструмент https://geosaero.ru/seedlingsnet на полях клиентов в 2024 г.

Выводы: В ходе работы была достигнута поставленная цель, заключающаяся в разработке точного автоматического метода для мониторинга грибных заболеваний и сорных растений в полевых условиях. Использование данного метода востребована индустрией для повышения эффективности плантационных работ.

Финансирование: Работа выполнена за счет финансирования Курчатовского геномного центра Федерального исследовательского центра ИЦиГ СО РАН, соглашение с Министерством образования и науки РФ № 075-15-2019-1662.

Image analysis-based methods for monitoring fungal diseases and weeds in the field

Genaev M.A.[1, 3], Kozhekin M.V.[1], Koval V.S.[1, 3], Afonnikov D.A.[1, 2, 3]*

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *ada@bionet.nsc.ru*

Key words: deep learning; computer vision; segmentation; detection

Motivation and Aim: One of the urgent problems in large agricultural holdings is monitoring of crops to identify diseases and centers of weed vegetation. Cereal diseases caused by pathogenic fungi can significantly reduce crop yields. Many crops are susceptible to them. The disease is difficult to control on a large scale; monitoring of crops, helps to identify diseases and weed centers at an early stage and take measures to prevent their spread. The aim of the work is to develop a method based on image analysis for monitoring fungal diseases and weeds in the field.

Methods and Algorithms: One of the effective methods of monitoring is the identification of diseases based on the analysis of digital images with the possibility of obtaining them in the field, using mobile devices. In this paper, a method for recognizing five fungal diseases of shoots in cereals (leaf rust, stem rust, yellow rust, powdery mildew and septoriosis) both individually and in multiple diseases, with the possibility of identifying the stage of plant development. A dataset of 2,414 images of fungal diseases of wheat

was generated, for which expert labeling by disease type was performed. More than 80 % of the images in the dataset correspond to single disease labels, 12 % are represented by healthy plants, and 6 % of the images are represented by multiple diseases. In the process of creating this dataset, a method to reduce the degeneracy of the training data based on an image hashing algorithm was applied.

In order to identify weed hotspots in the field, UAVs were used to acquire geospatial raster files. Computer vision methods were used to reconstruct the crop density map and determine the contours of cultivated and weed plants.

Results: A dataset of 2,414 images of fungal diseases of wheat was generated, for which expert labeling by disease type was performed. More than 80 % of the images in the dataset correspond to single disease labels, 12 % are represented by healthy plants, and 6 % of the images are represented by multiple diseases. In the process of creating this dataset, a method to reduce the degeneracy of training data based on image hashing algorithm was applied. The disease recognition algorithm is based on CNN with EfficientNet architecture. The best accuracy (0.942) was shown by the network with a learning strategy based on augmentation and transfer of image styles. The recognition method is implemented as a bot on the Telegram platform that allows users to evaluate plants from an image in the field. The dataset and model are publicly posted and available at http://wfd.sysbio.ru.

For field monitoring and using UAVs, we have developed a system SeedlingsNet for automatic estimation of the number of sprouts on RGB images of fields in GeoTIFF format with a resolution of 1–2 cm/pix. Identification of seedlings of wide-row field crops is carried out with an accuracy of more than 95 % compared to manual identification of seedlings. GeosAero plans to use the obtained tool https://geosaero.ru/seedlingsnet in the fields of its customers in 2024.

Funding: The research was funded by the Kurchatov Genomic Center of the Federal Research Center ICG SB RAS, agreement with the Ministry of Education and Science of the Russian Federation No. 075-15-2019-1662.

doi 10.18699/bgrs2024-06-06

Открытие новой ветви липоксигеназного каскада

Горина С.*, Ланцова Н., Ильина Т., Егорова А., Топоркова Я., Гречкин А.
Казанский институт биохимии и биофизики – обособленное структурное подразделение ФИЦ КазНЦ РАН, Казань, Россия
**gsvetlana87@gmail.com*

Ключевые слова: оксилипины; липоксигеназа; гидропероксидлиаза; *Cucumis sativus* L.

Мотивация и цель: Растительный липоксигеназный каскад является источником различных регуляторных оксилипинов, которые участвуют в передаче сигналов, адаптации к стрессовым факторам и защитных реакциях [1, 2]. Регио- и стереоспецифическая диоксигенация жирных кислот при участии липоксигеназ (ЛОГ) во многом определяет направления превращений гидроперекисей жирных кислот в различные оксилипины. Хорошо известно, что растительные ЛОГ окисляют линолевую и α-линоленовую кислоты до 9- и 13-гидроперекисей [2]. При изучении профилей оксилипинов растений огурца мы обнаружили значительное содержание 16-гидроперекиси α-линоленовой кислоты (16-ГПОТ), а также продуктов ее превращения, таких как 15-гидрокси-9,12-пентадекадиеновая кислота. Это соединение является продуктом восстановления 15-оксо-9,12-пентадекадиеновой кислоты, образованной при участии гидропероксидлиазы (ГПЛ).

ГПЛ принадлежит к неклассическим цитохромам Р450 семейства CYP74, которые, в отличие от большинства монооксигеназ Р450, для своей каталитической активности не нуждаются в молекулярном кислороде и окислительно-восстановительных партнерах. Субстраты – гидроперекиси полиненасыщенных жирных кислот – являются донорами кислорода и электронов. ГПЛ и ЛОГ образуют ГПЛ-ветвь ЛОГ пути, приводящую к образованию С6 и С9 альдегидов и спиртов, а также ω-оксокислот [3]. Эти соединения играют важную роль во взаимодействиях растение-растение и растение-травоядные, выполняя сигнальную роль в активации защитных систем растения, и участвуют непосредственно в защите растений, проявляя антимикробную и фунгицидную активность [4, 5]. Некоторые из этих соединений получили специальное название «летучие вещества зеленых листьев». Они вызывают характерный запах, который возникает, когда происходит повреждение листьев. Это ценные химические соединения, которые используются в пищевой и парфюмерной промышленности. Таким образом, целью нашей работы было выявление ферментов биосинтеза 16-ГПОТ и образующихся из нее оксилипинов.

Методы и алгоритмы: Открытые рамки считывания целевых генов были клонированы в векторах для экспрессии белков системы pET. Рекомбинантные белки очищали методами металлоаффинной и анионообменной хроматографии. Продукты (оксилипины), очищенные с помощью ВЭЖХ, идентифицировали с помощью газовой хромато-масс-спектрометрии (ГХ-МС) и ЯМР, включая ^{1}H-ЯМР, 2D-COSY, HSQC и HMBC.

Результаты: Профилирование эндогенных оксилипинов листьев, цветков и кожуры плодов огурца (*Cucumis sativus* L.) методом ГХ-МС выявило значительное

содержание 16-гидрокси-9,12,14-октадекатриеновой кислоты (16-ГОТ). При инкубации гомогенатов этих органов с α-линоленовой кислотой преобладающим продуктом была 16(*S*)-ГПОТ. Целенаправленный протеомный анализ этих тканей выявил наличие нескольких высокогомологичных изоформ предполагаемой «9*S*-ЛОГ тип 6». Один из этих изоферментов (CsLOX3, полипептид длиной 877 аминокислотных остатков) был получен в гетерологичной системе экспрессии с использованием клеток *E. coli* и проявлял 16(*S*)- и 13(*S*)-липоксигеназную активность по отношению к α-линоленовой и линолевой кислотам соответственно. При этом α-линоленовая кислота была предпочтительным субстратом. Молекулярные структуры 16(*S*)-ГОТ и 16(*S*)-ГПОТ (Me или Me/TMC) были подтверждены данными масс-спектров ^{1}H-ЯМР, 2D^{1}H-^{1}H-COSY, TOCSY, HMBC и HSQC, а также методом хирально-фазовой ВЭЖХ. Таким образом, липоксигеназа CsLOX3, катализирующая образование 16(*S*)-ГПОТ, является первой обнаруженной 16(*S*)-ЛОГ (ω3-ЛОГ) [6].

Для подтверждения предположения о катализируемом гидропероксидлиазой превращении 16(*S*)-ГПОТ, 16(*S*)-ГПОТ инкубировали с рекомбинантной ГПЛ огурца (CYP74B6). CYP74B6 обладал высокой активностью по отношению к 16-ГПОТ. Продукт восстановления $NaBH_4$ (Me/TMC) был идентифицирован как 15-гидрокси-9,12-пентадекадиеновая кислота. Таким образом, рекомбинантная ГПЛ огурца (CYP74B6) обладает беспрецедентной активностью 16-ГПЛ, расщепляя 16-ГПОТ на фрагмент C15, 15-оксо-9,12-пентадекадиеновую кислоту и комплементарный летучий фрагмент C3, пропионовый альдегид. Кроме того, эти соединения были выявлены и при профилировании оксилипинов листьев и кожуры плодов огурца.

Выводы: Нами обнаружена новая ветвь метаболизма α-линоленовой кислоты, а именно 16-ЛОГ/ГПЛ ветвь. Растения огурца, а также рекомбинантная гидропероксидлиаза CYP74B6 огурца обладали беспрецедентной 16-ГПЛ активностью. 16-ЛОГ/ГПЛ путь биосинтеза оксилипинов представляет собой новую ветвь растительного липоксигеназного каскада.

Финансирование: Биоинформатический анализ проводился при финансовой поддержке государственного задания Федерального исследовательского центра «КазНЦ РАН». Исследования рекомбинантных ферментов поддержаны грантами РНФ 23-14-00350 (ГПЛ) и 24-14-00418 (ЛОГ).

Discovery of 16(*S*)-lipoxygenase/16-hydroperoxide lyase pathway in green tissues of cucumber (*Cucumis sativus* L.) plants

Gorina S.*, Lantsova N., Iljina T., Egorova A., Toporkova Y., Grechkin A.

Kazan Institute of Biochemistry and Biophysics – Subdivision of the Federal State Budgetary Institution of Science "Kazan Scientific Center of the Russian Academy of Sciences", Kazan, Russia

**gsvetlana87@gmail.com*

Key words: oxylipins; lipoxygenase; hydroperoxide lyase; *Cucumis sativus* L.

Motivation and Aim: The plant lipoxygenase cascade is a source of various regulatory oxylipins that play a role in cell signaling, stress adaptation, and immune response [1, 2]. Regio- and stereospecific dioxygenation of fatty acids by lipoxygenases (LOXs) largely

determine the directions of conversion of fatty acid hydroperoxides to various oxylipins. Plant LOXs are well known to oxidize linoleic and α-linolenic acids to 9- and 13-hydroperoxides [2]. When studying the oxylipin profiles in cucumber plants, we found a significant content of 16-hydroperoxide of α-linolenic acid (16-HPOT), as well as its conversion products, such as 15-hydroxy-9,12-pentadecadienoic acid. This compound is a product of the reduction of 15-oxo-9,12-pentadecadienoic acid, formed by hydroperoxide lyase (HPL).

HPL belongs to the non-classical cytochromes P450 of the CYP74 family, which, unlike most P450 monooxygenases, do not require molecular oxygen and redox partners for their catalytic activities. Substrates, hydroperoxides of polyunsaturated fatty acids, act as donors of oxygen and electrons. HPL and LOX form the HPL-branch of the LOX pathway, leading to the formation of C6 and C9 aldehydes and alcohols, as well as ω-oxoacids [3]. These compounds play important roles in plant-plant and plant-herbivore interactions, function as signals activating systemic defense, and are involved directly in plant defense as antimicrobials and fungicides [4, 5]. A number of these compounds are specifically named Green Leaf Volatiles (GLVs). They cause a characteristic odor that occurs when leaves are crushed or otherwise injured. They are valuable chemicals that are used in the food and perfume industries.

Thus, the purpose of our work was to identify enzymes for the biosynthesis of 16-HPOT and the oxylipins formed from it.

Methods and Algorithms: The open reading frames of target genes were cloned into expression vectors of the pET system. Recombinant proteins were purified by immobilized metal-affinity and anion-exchange chromatography. Products (oxylipins) purified by HPLC were identified by GC-MS and NMR data, including ^{1}H-NMR, 2D-COSY, HSQC, and HMBC.

Results: The GC-MS profiling of the endogenous oxylipins (Me/TMS) from cucumber (*Cucumis sativus* L.) leaves, flowers, and fruit peels revealed a remarkable abundance of 16-hydroxy-9,12,14-octadecatrienoic acid (16-HOT). Incubations of homogenates from these organs with α-linolenic acid yielded 16(*S*)-hydroperoxide (16-HPOT) as a predominant product. Targeted proteomic analyses of these tissues revealed the presence of several highly homologous isoforms of the putative "9*S*-lipoxygenase type 6". One of these isoenzymes (CsLOX3, an 877 amino acid polypeptide) was prepared by heterologous expression in *E. coli* and exhibited 16(*S*)- and 13(*S*)-lipoxygenase activities toward α-linolenic and linoleic acids, respectively. Furthermore, α-linolenate was a preferred substrate. The molecular structures of 16(*S*)-HOT and 16(*S*)-HPOT (Me or Me/TMS) were unequivocally confirmed by the mass spectral data, ^{1}H-NMR, 2D ^{1}H-^{1}H-COSY, TOCSY, HMBC, and HSQC spectra, as well as enantiomeric HPLC analyses. Thus, the vegetative CsLOX3, biosynthesizing 16(*S*)-HPOT, is the first 16(*S*)-LOX and ω3-LOX ever discovered [6].

To confirm the proposed HPL chain cleavage, the 16(*S*)-HPOT was prepared and incubated with the recombinant cucumber HPL CYP74B6 enzyme. The CYP74B6 possessed high activity towards 16-HPOT. The $NaBH_4$-reduced product (Me/TMS) was identified as 15-hydroxy-9,12-pentadecadienoic acid. Thus, the cucumber plants, as well as the recombinant cucumber HPL CYP74B6, possessed unprecedented 16-HPL activity, cleaving 16-HPOT into a C15 fragment, 15-oxo-9,12-pentadecadienoic acid, and a complementary volatile C3 fragment, propionic aldehyde.

Conclusion: We have discovered a new branch of metabolism in the transformation of α-linolenic acid, namely the 16-LOX/HPL branch. The cucumber plants as well as the

recombinant cucumber HPL CYP74B6 possessed unprecedented 16-HPL activity. The 16-LOX/16-HPL route of oxylipin biosynthesis presents a novel facet of the plant LOX pathway.

Funding: Bioinformatic analysis was carried out with financial support from the state assignment of the Federal Research Center "KazSC of RAS". Studies of the recombinant enzymes were supported by grants 23-14-00350 (HPL) and 24-14-00418 (LOX) from the Russian Science Foundation.

Список литературы/References

1. Mosblech A., Feussner I., Heilmann I. Oxylipins: Structurally diverse metabolites from fatty acid oxidation. *Plant Physiol. Biochem*. 2009;47:511-517. doi 10.1016/j.plaphy.2008.12.011
2. Wasternack C., Feussner I. The Oxylipin Pathways: Biochemistry and Function. *Annu. Rev. Plant Biol.* 2018;69:363-386. doi 10.1146/annurev-arplant-042817-040440
3. Grechkin A.N. Hydroperoxide lyase and divinyl ether synthase. *Prostaglandins Other Lipid Mediat.* 2002;68-69:457-470. doi 10.1016/s0090-6980(02)00048-5
4. Matsui K. Green leaf volatiles: Hydroperoxide lyase pathway of oxylipin metabolism. *Curr. Opin. Plant Biol*. 2006;9:274-280. doi 10.1016/j.pbi.2006.03.002
5. Frost C.J., Mescher M.C., Dervinis C., Davis J.M., Carlson J.E., De Moraes C.M. Priming defense genes and metabolites in hybrid poplar by the green leaf volatile cis-3-hexenyl acetate. *New Phytol.* 2008;180:722-734. doi 10.1111/j.1469-8137.2008.02599.x
6. Gorina S.S., Egorova A.M., Lantsova N.V., Toporkova Y.Y., Grechkin A.N. Discovery of α-Linolenic Acid 16(S)-Lipoxygenase: Cucumber (Cucumis sativus L.) Vegetative Lipoxygenase 3. *J. Mol. Sci.* 2023;24(16):12977. doi 10.3390/ijms241612977

doi 10.18699/bgrs2024-06-07

Восстановление функции гена *HvMyc2*, контролирующего голубую окраску зерна ячменя, путем его направленной модификации

Зыкова Т.[1, 2]*, Егорова А.[1, 2, 3], Шоева О.[1, 2, 3], Хертиг К.[4], Коэппель И.[4], Хикель Ш.[4], Короткова А.[2, 3], Кумлен Й.[4], Герасимова С.[2, 3], Хлесткина Е.[1, 2, 3]

[1] *Новосибирский государственный университет, Новосибирск, Россия*

[2] *Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР), Санкт-Петербург, Россия*

[3] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[4] *Институт генетики растений и исследований сельскохозяйственных культур им. Лейбница, Гатерслебен, Германия*

* *t.zykova@g.nsu.ru*

Ключевые слова: *Hordeum vulgare*; антоцианы; алейроновый слой; повышение пищевой ценности

Мотивация и цель: В настоящее время актуальны исследования генетической основы метаболических реакций культурных растений, связанных с образованием биоактивных соединений с антиоксидантными свойствами – флавоноидов, и в том числе антоцианов. Положительное влияние данных биоактивных веществ на липидный обмен и защиту от окислительного стресса у модельных животных и человека приводит к снижению риска развития ряда заболеваний [1]. У злаков образование антоциановых пигментов происходит в различных слоях оболочки зерновок. Так, алейроновый слой, прилежащий к тканям эндосперма, может накапливать антоциан дельфинидин, в результате чего зерно ячменя (*Hordeum vulgare* L.) приобретает голубой оттенок.

Биосинтез флавоноидов является результатом работы структурных генов, кодирующих ферменты, и регуляторных генов, кодирующих, например, транскрипционные факторы. Существуют данные о функциях и последовательностях структурных генов биосинтеза антоцианов, но для полного описания механизма его контроля необходимо выяснить роль регуляторных факторов. Это позволит создавать растения с зернами, имеющими повышенное содержание антоцианов.

Ранее было показано, что ген *HvMyc2* кодирует транскрипционный фактор, активирующий работу генов ферментов биосинтеза антоцианов в алейроновом слое зерен ячменя [2]. Была также обнаружена корреляция между фенотипом неокрашенных зерен ячменя и однонуклеотидной инсерцией в данном гене, вызывающей сдвиг рамки считывания с образованием раннего стоп-кодона.

Целью работы стало уточнение функции гена *HvMyc2* в регуляции биосинтеза антоцианов зерновок ячменя путем восстановления рамки считывания данного гена в неокрашенном сорте посредством внесения в целевой участок небольших делеций при помощи РНК-направленной эндонуклеазы Cas9.

Методы и алгоритмы: Мутантные растения были получены в результате агробактериальной трансформации незрелых зародышей сорта Golden Promise, имеющего неокрашенные зерновки. Затем после последовательных

самоопылений в теплице получали поколения T1–T5. Путем генотипирования поколений T0–T4 с помощью глубокого секвенирования и метода Сэнгера были отобраны гомозиготные мутантные линии. Дополнительно было проведено тестирование на наличие вставок трансгена Cas9. Анализ фенотипа зерен полученных линий был проведен с помощью световой микроскопии, биохимического определения антоцианов [3] и определения уровня экспрессии генов ферментов биосинтеза данных соединений.

Световая микроскопия была использована для визуальной оценки количества антоциановых пигментов в алейроновом слое отдельных зерен по четырехбалльной шкале. Для этого с растения собирали три колоса, из каждого выбирали по 10 зерен. В результате было проанализировано 30 зерен с растения и по 90 зерен каждого генотипа.

Оценка экспрессии структурных генов биосинтеза антоцианов флаванон-3-гидроксилазы (*F3h*), флавоноид-3'-гидроксилазы (*F3'h*), флавоноид-3',5'-гидроксилазы (*F3'5'h*) и антоцианидинсинтазы (*Ans*) была проведена с помощью полимеразной цепной реакции в реальном времени с красителем SYBR (Applied Biosystems, США). В качестве референсных использовали гены *Actin* и *Mdh*, кодирующие актин и малатдегидрогеназу соответственно. Результаты расчетов были проанализированы с помощью теста Краскела–Уоллиса и последующих попарных сравнений Манна–Уитни с поправкой Бонферрони на множественное тестирование (Past, версия 4.03).

Результаты: Были созданы две конструкции, включающие ген Cas9 и направляющую РНК к одному из двух целевых участков в кодирующей части гена *HvMyc2*. После тестирования активности конструкций на протопластах ячменя была проведена агробактериальная трансформация незрелых зародышей сорта ячменя Golden Promise. Среди полученных растений поколения T0 49 содержали мутации в целевом гене, для дальнейшего отбора в поколениях T1–T4 выбрали восемь растений. В поколении T4 две линии имели *HvMyc2* дикого типа, две другие – содержали делецию 4 п. н. в первом или втором целевом сайте, одна – делецию 11 п. н., а другие четыре – делецию 1 п. н.

В линиях с делециями 1 и 4 п. н. функция целевого гена была восстановлена, в результате чего произошла активация биосинтеза антоцианов в алейроновом слое зерновки. Накопление целевых соединений в зерне было подтверждено с помощью световой стереомикроскопии и количественного определения содержания антоцианов. Анализ экспрессии генов биосинтеза антоцианов показал, что восстановление работы гена *HvMyc2* приводит к увеличению уровня транскриптов структурных генов, вовлеченных в общие для всех антоцианов реакции.

Выводы: Таким образом, данное исследование продемонстрировало возможность использования технологии генного редактирования для улучшения питательной ценности зерен ячменя путем увеличения содержания антоцианов.

Финансирование: Исследование поддержано грантом РНФ (№ 21-66-00012).

Restoring the function of the *HvMyc2* gene, which controls the blue color of barley grain, through its targeted modification

Zykova T.[1, 2]*, Egorova A.[1, 2, 3], Shoeva O.[1, 2, 3], Hertig K.[4], Koeppel I.[4], Hikel S.[4], Korotkova A.[2, 3], Kumlehn J.[4], Gerasimova S.[2, 3], Khlestkina E.[1, 2, 3]

[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia*
[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[4] *Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany*
* *t.zykova@g.nsu.ru*

Key words: *Hordeum vulgare*; anthocyanins; aleurone layer; increase in nutritional value

Motivation and Aim: Currently, research into the genetic basis of metabolic reactions of cultivated plants associated with the formation of bioactive compounds with antioxidant properties – flavonoids, including anthocyanins – is relevant. The positive effect of these bioactive substances on lipid metabolism and protection from oxidative stress in model animals and humans leads to a reduction in the risk of developing a number of diseases [1]. In cereals, the formation of anthocyanin pigments occurs in various layers of the grain hull. Consequently, the aleurone layer in close proximity to the endosperm tissues can accumulate the anthocyanin delphinidin, thereby imparting a blue hue to barley grains (*Hordeum vulgare* L.).

The biosynthesis of flavonoids is the result of the combined action of structural enzyme genes and regulatory genes, which encode, for instance, transcription factors. There is data on the functions and sequences of structural genes for anthocyanin biosynthesis, but to fully describe the mechanism of its control, it is necessary to clarify the role of regulatory factors. This will enable the creation of plants with grains that have a high content of anthocyanins.

It was previously shown that the *HvMyc2* gene encodes a transcription factor that activates the genes for anthocyanin biosynthesis enzymes in the aleurone layer of barley grains [2]. A correlation was also identified between the phenotype of uncolored barley grains and a single nucleotide insertion in this gene, which resulted in a frameshift and the formation of an early stop codon.

The objective of this study was to elucidate the function of the *HvMyc2* gene in the regulation of anthocyanin biosynthesis in barley grains by restoring the reading frame of this gene in an uncolored variety through the introduction of small deletions into the target region using RNA-directed endonuclease Cas9.

Methods and Algorithms: The mutant plants were obtained as a result of agrobacterial transformation of immature embryos of the Golden Promise variety, which has uncolored grains. Subsequently, following a series of self-pollinations conducted within a greenhouse setting, five generations (T1–T5) were generated. Genotyping of generations T0–T4 was conducted using deep sequencing and the Sanger method, with the aim of identifying lines homozygous for mutations. Additionally, testing was performed to detect the presence of Cas9 transgene insertions. The phenotype of the grains of the obtained lines was analyzed using light microscopy, biochemical determination of anthocyanins [3], and assessment of the level of expression of structural genes of the biosynthesis of these compounds.

Light microscopy was employed to assess the quantity of anthocyanin aggregates in the aleurone layer of individual grains on a four-point scale. To this end, three spikes were harvested from the plant, and 10 grains were selected from each spike. Consequently, 30 grains per plant and 90 grains of each genotype were subjected to analysis.

The expression of the structural genes of anthocyanin biosynthesis, namely flavanone 3-hydroxylase (*F3h*), flavonoid 3'-hydroxylase (*F3'h*), flavonoid 3',5'-hydroxylase

(*F3'5'h*) and anthocyanidin synthase (*Ans*), was evaluated using real-time polymerase chain reaction with SYBR dye (Applied Biosystems, USA). The *Actin* and *Mdh* genes, which encode actin and malate dehydrogenase, respectively, were employed as reference genes. The results of the calculations were subjected to analysis using the Kruskal–Wallis test and subsequent pairwise Mann–Whitney comparisons with Bonferroni correction for multiple testing (Past, version 4.03).

Results: Two constructs were designed, including the Cas9 gene and a guide RNA to one of two target regions in the coding part of the *HvMyc2* gene. Following the assessment of construct activity in barley protoplasts, agrobacterial transformation of immature embryos of the barley variety Golden Promise was conducted. Among the plants of the T0 generation, 49 exhibited mutations in the target gene. Eight of these plants were selected for further selection in the T1–T4 generations. In the T4 generation, two lines exhibited wild-type *HvMyc2*, while the other two displayed a 4 bp deletion at the first or second target site, one exhibited an 11 bp deletion, while the other four lines exhibited a 1 bp deletion.

In lines exhibiting deletions of 1 and 4 bp, the function of the target gene was restored, resulting in the activation of anthocyanin biosynthesis in the aleurone layer of the grain. The accumulation of target compounds in the grain was confirmed using light microscopy and quantification of anthocyanin content. The expression of anthocyanin biosynthesis genes was analyzed, and it was found that the restoration of the *HvMyc2* gene resulted in an increase in the level of transcripts of structural genes involved in reactions that are common to all anthocyanins.

Conclusion: This study demonstrated the feasibility of using gene editing technology to enhance the nutritional value of barley grains by increasing the anthocyanin content.

Funding: The study is supported by a grant from the Russian Science Foundation (No. 21-66-00012).

Список литературы/References

1. Lin B.W. et al. Effects of anthocyanins on the prevention and treatment of cancer. *Br J Pharmacol*. 2017;174(11):1226-1243
2. Strygina K.V., Börner A., Khlestkina E.K. Identification and characterization of regulatory network components for anthocyanin synthesis in barley aleurone. *BMC Plant Biol*. 2017;17(Suppl. 1):184. doi 10.1186/s12870-017-1122-3
3. Abdel-Aal E.S.M., Hucl P. A rapid method for quantifying total anthocyanins in blue aleurone and purple pericarp wheats. *Cereal Chem*. 1999;76(3):350-354

doi 10.18699/bgrs2024-06-08

Экогеномная стресс-устойчивость растений как стратегия и тактика супрамолекулярно-протеомной системы биологии развития

Иванова Э.А.
Уфимский институт биологии, УФИЦ РАН, Уфа, Россия
** fiona_belobor@mail.ru*

Ключевые слова: протеомика; интерфазная топология хроматина; супрамолекулярная биохимия; кариогеномика пшеницы; сигнальные системы

Мотивация и цель: В мире растений многие системные проблемы, выполняемые *“Plant Gen”* из области экологии, энергетики, передачи информации и т. д., рассматриваются в качестве модельных систем, так как они эффективны, экологически безупречны и одновременно просты. Человек учится у растений, постигает их способы и методы в решении уже своих проблем.
Молекулярная био-физико-химия, занимающаяся строением и свойствами биологически функциональных молекул и атомно-молекулярным истолкованием явлений жизнедеятельности, за короткое время стала широко развитой наукой, активно делающей шаги навстречу возрастающей сложности изучения запрограммированных супрамолекулярных систем, вплоть до экосистем.
Методы и алгоритмы: Идеальным модельным объектом в этом случае являются коллекционные семена ВИР, преобразованные селекционерами из яровой пшеницы в озимую и из последней вновь в яровую.
Экспериментальный биохимический подход заключался: в выделении из клеточных ядер тотальной хроматиновой матрицы (ТХМ) *супра*-структур: Нп-лабильного нуклеоплазменного хроматина, Хр-I, Хр-II хроматинов, соответственно непрочно- и прочно связанных с ядерным матриксом (ЯМ), и самого ЯМ, из которых были выделены *супер*молекулярные ансамбли: «линкерных», «коровых» гистонов и негистонов.
Результаты: Показан алгоритм особенностей биологической специфичности морфогенеза и структурной устойчивости геномно-протеомной основы ТХМ модельной системы, коллекционных зародышей семян пшеницы в процессе их органоспецифического, координированно-закономерного роста при переключении подпрограмм развития, где происходит изменение протеомного позиционирования в супермолекулярных ансамблях: линкерных, «коровых» и «негистоновых» белков ТХМ разных генетических подсистем (мезокотиль, корень, высокодифференцированный зародыш) соответственно: донор (яровой)→переведен в озимый (донор озимый-фенотип)→переведен вновь в яровой-фенотип.
Выводы: На основании распределения нуклеосомного аргинин-богатого «корового» гистона (Н3-Н4)’’ на поверхности раздела ТХМ: донора (яровой) Нп=ХрI (мезокотиль)→переведенного в озимый (донор озимый-фенотип) Нп>ХрII≥ЯМ (корень)→переведенного вновь в яровой-фенотип Нп>ХрI>ЯМ>ХрII (высокодифференцированный зародыш); предполагается

возможное переключение генетических подпрограмм развития в генетических подсистемах целостного организма, которое осуществляется за счет комбинаторного принципа протеомных ансамблей, потенциальных эпигенетических сетей «гистонового кода», в условиях экосистемы окружающей среды.

Финансирование: Работа выполнена в рамках государственного задания Минобрнауки России № 075-00326-19-00 по теме № НИОКТР АААА-А18-118022190104-7. В работе использована приборная база Центра коллективного пользования «Агидель» УФИЦ РАН.

Eco-genomic stress-resistance of plants as strategy and tactics of the supramolecular-proteomic system of developmental biology

Ivanova E.A.

Ufa Institute of Biology, Ufa Federal Research Centre, RAS, Ufa, Russia

** fiona_belobor@mail.ru*

Key words: proteomics; interphase chromatin topology; supramolecular biochemistry; karyogenomics wheat; signaling systems

Motivation and Aim: In the world of plants, many systemic *"Plant Gen"* problems solve in the field of ecology, energy, information transmission, and so on, are considered as model systems, since they are effective, environmentally friendly and at the same time simple. A person learns from plants, comprehends their ways and methods in solving their own problems.

Molecular bio-physicochemistry, which deals with the structure and properties of biologically functional molecules and the atomic-molecular interpretation of life phenomena, has in a short time become a widely developed science, actively taking steps towards the increasing complexity of studying programmed supramolecular systems, up to ecosystems.

Methods and Algorithms: The ideal model object, in this case, is the collection seeds of VIR, transformed by breeders from spring wheat to winter wheat and from the latter again to spring wheat.

The experimental biochemical approach consisted of isolating supra-structures from the cell nuclei of the total chromatin matrix (TChM): Hn-labile nucleoplasmic chromatin, Chr-I, Chr-II chromatins, respectively, loosely and firmly associated with the nuclear matrix (NM) and the NM itself, from of which supermolecular ensembles were isolated: "linker", "core" histones and non-histones.

Results: An algorithm for the features of the biological specificity of morphogenesis and structural stability of the genetic and proteomic basis of the TChrM model system, collection germs of wheat seeds, in the process of their organ-specific, coordinated-regular growth when switching development subprograms is shown where an experimental analysis of proteomic positioning in *super*molecular assemblies was carried out: "linker", "core" and "non-histone" proteins in different genetic subsystems (mesocotyl → root → highly differentiated embryo), respectively: donor (spring) → transferred to winter (donor winter-phenotype) → transferred back to spring-phenotype.

Conclusion: Based on the distribution of nucleosomal arginine-rich "core" histone (H3-H4)" on the TChM interface: donor (spring) Np=ChrI (mesocotyl) → transferred to

winter (donor winter-phenotype) Np>ChrII≥NM (root)→transferred again into the spring phenotype Np>ChrI>NM>ChrII) (highly differentiated embryo); possible switching of genetic subroutines of development in the genetic subsystems of the whole organism is assumed, which is carried out due to the combinatorial principle of proteomic ensembles, potential epigenetic networks of the "histone code", in the conditions of environmental ecosystems.

Funding: The work was carried out within the framework of the state assignment of the Ministry of Education and Science of Russia No. 075-00326-19-00 on the topic No. NIOKTR AAAA-A18-118022190104-7. The instrument base of the Center for Collective Use "Agidel" of the UFRC RAS was used in the work.

doi 10.18699/bgrs2024-06-09

WheatHeadDetection. Модернизация моделей обнаружения колосьев и протокола сбора данных

Кожекин М.[1, 2]*, Афонников Д.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

** kozhekinMV@bionet.nsc.ru*

Ключевые слова: феномика; полевая феномика; компьютерное зрение; пшеница; БПЛА; тележка

Мотивация и цели: Пшеница – одна из наиболее важных сельскохозяйственных культур, которая обеспечивает питанием существенную часть населения в мире. Для увеличения эффективности проведения селекционно-генетических экспериментов необходима разработка информационных технологий фенотипирования растений.

В работе [1] мы представили систему компьютерного зрения для подсчета колосьев пшеницы с использованием БПЛА DJI Mavic 2 и Phantom 4. Значения коэффициентов корреляции между плотностью колосьев, предсказанной системой, и продуктивным числом стеблей оказались недостаточными для практического применения. Это обусловлено низкой точностью подсчета колосьев и недостатками протокола съемки.

В работе [2] произведена оценка решений победителей соревнований Global Wheat Head Detection на новых данных. Они показывали низкую точность. Использование повернутых ограничивающих прямоугольников могло бы исправить ситуацию.

В используемом протоколе съемки были две проблемы. Во-первых, облет коптером проводился поперек делянок. Постоянный набор высоты вкупе с различной высотой сортов пшеницы приводил к росту дисперсии размера колосьев на снимках. Во-вторых, используемый коптер не имел системы коррекции координат, что снижало точность параметра высоты полета и GPS координат.

Необходимость иметь альтернативный инструмент для сбора данных, а также желание оценить теоретический потенциал моделей детекции колосьев привели к созданию наземной платформы для сбора данных в виде тележки. Поля, на которых проводятся сортоиспытания, находятся каждый раз на новом месте, поэтому было выдвинуто требование к мобильности платформы.

Цель работы заключается в улучшении методов обнаружения колосьев и поиске оптимального метода сбора данных.

Методы и алгоритмы: Используя данные из [3] и снимки, размеченные ранее, были обучены несколько моделей с повернутыми ограничивающими прямоугольниками.

Чтобы определить теоретическую производительность моделей детекции и иметь запасной инструмент для сбора данных, была собрана доступная и мобильная наземная платформа из алюминиевого профиля 20×20×1000 мм и колес от тачки.

Результаты: В результате точность моделей, учитывающих направление колоса, оказалась значительно выше.

Модернизированый протокол съемки позволяет исправить проекцию данных на карту. На ортофотопланах, собранных нашей программой, стали различимы междурядья.
С использованием тележки был получен набор данных одного яруса. Каждая делянка была снята в трех точках.
Выводы: Применение моделей детекции, учитывающих геометрию колоса, повысило точность их обнаружения. Модернизированный протокол съемки повысил качество отображения данных на карту. Наземная платформа позволила оценить влияние дисперсии параметра высоты и наличие вибрации при съемке.
Финансирование: Работа выполнена за счет финансирования Курчатовского геномного центра ФИЦ ИЦиГ СО РАН, соглашение с Министерством образования и науки РФ № 075-15-2019-1662.

WheatHeadDetection. Modernisation of wheat head detection models and data collection protocol

Kozhekin M.[1, 2]*, Afonnikov D.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

** kozhekinMV@bionet.nsc.ru*

Key words: phenomics; field phenomics; computer vision; wheat; UAV; cart

Motivation and Aim: Wheat is one of the most important crops that feeds a significant portion of the world's population. In order to increase the efficiency of plant breeding and genetic experiments, it is necessary to develop information technologies for plant phenotyping.
In [1], we presented a computer vision system for counting wheat ears using DJI Mavic 2 and Phantom 4 UAVs. The values of the correlation coefficients between the ear density predicted by the system and the productive number of stems were insufficient for practical applications. This is due to the low accuracy of spike counting and shortcomings of the imaging protocol.
It is shown in [2] that the solutions of Global Wheat Head Detection competitions winners show low accuracy on new data. The use of oriented bounding boxes could fix the problem.
There were 2 flaws in the data acquisition protocol. Firstly, the copter flew across the plots. Constant altitude gain coupled with different heights of wheat varieties resulted in increased dispersion of ear size on the images. Secondly, the copter used did not have a RTK system, which reduced the accuracy of the altitude parameter and GPS coordinates.
Methods and Algorithms: With the use of data from [3] and previously labeled images, several models with rotated bounding boxes were trained. UAV telemetry has been added to the metadata from the images: the relative flight altitude of the GPS receiver has been replaced by ultrasonic sensors data. The flight mission planner algorithm was changed: the flight direction parallel to the plots was replaced. To determine the theoretical performance of detection models and to have a spare tool for data collection, an accessible and mobile ground platform was assembled from a 20×20×1000 mm aluminum profile and wheelbarrow wheels.

Results: As a result, the accuracy of models taking into account the direction of the ear turned out to be significantly higher. The modernized data acquisition protocol allows you to correct the projection of data onto the map. On the orthophotomaps collected by our program, row spacing became visible. Using a cart, a single tier data set was obtained. Each plot was filmed at 3 points.
Conclusions: The use of detection models that take into account the geometry of the ear has increased the accuracy of their detection. The modernized survey protocol has improved the quality of data display on the map. The ground platform made it possible to evaluate the influence of the dispersion of the height parameter and the presence of vibration during image capture.
Funding: This work was funded by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation No. 075-15-2019-1662.

Список литературы/References

1. Kozhekin M. et al. Wheat yield estimation based on analysis of UAV images at low altitude. *BIO Web Conf*. 2022;47:05006. doi 10.1051/bioconf/20224705006
2. David E. et al. Global wheat head detection challenges: Winning models and application for head counting. *Plant Phenomics.* 2023;5:0059. doi 10.34133/plantphenomics.0059
3. Badhon M.A., Stavness I. Fast rotated bounding box annotations for object detection. In: Agriculture-Centric Computation: First International Conference, ICA 2023, Chandigarh, India. Springer Nature, 2023;99-116. doi 10.1007/978-3-031-43605-5_8

doi 10.18699/bgrs2024-06-10

Модернизация моделей обнаружения колосьев на изображениях полевых посевов пшеницы

Кожекин М.[1, 2]*, Коваль В.[1, 2], Генаев М.[1, 2], Афонников Д.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

** kozhekinMV@bionet.nsc.ru*

Ключевые слова: феномика; полевая феномика; компьютерное зрение; пшеница; обнаружение и подсчет колосьев; колосья; БПЛА

Мотивация и цели: Пшеница – одна из наиболее важных сельскохозяйственных культур, которая обеспечивает питанием существенную часть населения в мире. Для увеличения эффективности проведения селекционно-генетических экспериментов необходима разработка информационных технологий фенотипирования растений. Одно из актуальных применений этих технологий – оценка урожайности посевов на основании анализа изображений и подсчета на них колосьев. Ранее мы представили систему компьютерного зрения для подсчета колосьев пшеницы в полевых посевах с использованием БПЛА DJI Mavic 2 и Phantom 4 [1]. Изображения посевов пшеницы, полученные с БПЛА, обрабатывались с помощью алгоритмов глубокого машинного обучения (нейронные сети архитектур Faster RCNN и Efficient-Det). Оценки количества колосьев на единицу площади посева, полученные с БПЛА и вручную, имели высокие значения коэффициентов корреляции (>0.54). Одно из возможных направлений увеличения точности – совершенствование алгоритмов распознавания колосьев на изображении [2]. В основе лежат алгоритмы детекции колосьев заданными прямоугольными рамками на изображении. Однако на изображениях посевов колосья располагаются в произвольных направлениях, что затрудняет их распознавание.

Методы и алгоритмы: Для распознавания колосьев на изображениях мы использовали алгоритм глубокого машинного обучения, основанный на использовании развернутых ограничивающих прямоугольников [3]. Использовались архитектуры сети Rotated Faster-RCNN, Rotated RetinaNet, ReDet (для развернутых прямоугольников) и EfficientDet (для вертикальных прямоугольников). Для обучения и тестирования мы использовали изображения посевов пшеницы из предыдущей работы [1].

Результаты: Показано, что использование развернутых прямоугольников для обучения позволяет улучшить точность подсчета колосьев на полевых изображениях: коэффициент детерминации увеличивается в среднем от 0.540 до 0.824.

Выводы: Применение моделей детекции, учитывающих геометрию колоса повысило точность их обнаружения.

Финансирование: Работа выполнена за счет финансирования Курчатовского геномного центра ФИЦ ИЦиГ СО РАН, соглашение с Министерством образования и науки РФ № 075-15-2019-1662.

Благодарности: Вычисления проводились с использованием ресурсов ЦКП «Биоинформатика».

Modernization of spike detection models for in-field wheat images

Kozhekin M.[1, 2]*, Koval V.[1, 2], Genaev M.[1, 2], Afonnikov D.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

**kozhekinMV@bionet.nsc.ru*

Key words: phenomics; field phenomics; computer vision; wheat; wheat spike detection and counting; wheat spike; UAV

Motivation and Aim: Wheat is one of the most important crops that feeds a significant portion of the world's population. In order to increase the efficiency of plant breeding and genetic experiments, it is necessary to develop information technologies for plant phenotyping. One of the current applications of these technologies is the assessment of crop yields based on image analysis and counting of spikes on them. Previously, we presented a computer vision system for counting wheat spikes in field crops using DJI Mavic 2 and Phantom 4 UAVs [1]. Images of wheat crops obtained from UAVs were processed using deep machine learning algorithms (neural networks of Faster RCNN and Efficient-Det architectures). Estimates of the number of spikes per unit of crop area obtained from a UAV and manually had high correlation coefficients (>0.54). One of the possible directions for increasing accuracy: improving detection algorithms for wheat spikes at image [2]. It is based on algorithms for detecting spikes, defined by axis-aligned bounding boxes in the image. However, in images of crops, the spikes are oriented in random directions, which makes them difficult to recognize.

Methods and Algorithms: To detect wheat spikes on images, we used a deep learning algorithm based on the use of rotated bounding boxes [3]. The network architectures used were Rotated Faster-RCNN, Rotated RetinaNet, ReDet (for rotated bounding boxes) and EfficientDet (for axis-aligned bounding boxes). For training and testing, we used images of wheat crops from previous work [1].

Results: It is shown that the use of rotated bounding boxes can improve the accuracy of counting spikes on field images: the coefficient of determination increases on average from 0.540 to 0.824.

Conclusions: The use of detection models that take into account the geometry of the spikes has increased the accuracy of their detection.

Funding: This work was funded by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation No. 075-15-2019-1662.

Acknowledgements: The calculations were performed using the resources of the Bioinformatics Center.

Список литературы/References

1. Kozhekin M. et al. Wheat yield estimation based on analysis of UAV images at low altitude. *BIO Web Conf.* 2022;47:05006
2. David E. et al. Global wheat head detection challenges: Winning models and application for head counting. *Plant Phenomics*. 2023;5:0059
3. Badhon M. A., Stavness I. Fast Rotated Bounding Box Annotations for Object Detection. In: Int. Conf. on Agriculture-Centric Computation. Cham: Springer Nature Switzerland, 2023;99-115

doi 10.18699/bgrs2024-06-12

Обновление приложения SeedCounter – оценка цветовых и текстурных характеристик зерен злаковых

Комышев Е.Г.[1]*, Генаев М.А.[1, 2], Афонников Д.А.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

** komyshev@bionet.nsc.ru*

Ключевые слова: мобильное приложение; цветовые характеристики; текстурные характеристики; зерна злаковых

Мотивация и цель: Оболочка зерна – это основной барьер между зерном и внешней средой, поэтому с ее характеристиками связан ряд важных биологических функций: поглощение влаги, жизнеспособность зерна, устойчивость к предуборочному прорастанию [1]. Цвет оболочки зерен злаков – важный признак, он характеризует пигменты и метаболиты, содержащиеся в ней, такие как каротиноиды, антоцианы, меланины. Помимо цвета, интерес представляет также текстура зерна. Например, шершавость оболочки зерен у пшеницы – это признак, который связан с восприимчивостью зерен поражению вредителями [2], технологическими свойствами зерна [3], содержанием влаги [4]. При изучении механизмов генетического контроля тех или иных признаков зерна селекционеры и генетики сталкиваются с необходимостью оценки цветовых и текстурных характеристик их оболочки.

Мобильные устройства все чаще стали использоваться в различных сферах научных наблюдений, в том числе и в агробиотехнологиях. Благодаря совершенствованию цифровой техники, качество цветопередачи на современных моделях камер становится все более точным. Это позволяет с высоким качеством оценивать цветовые характеристики органов растений. Разработанное нами ранее приложение SeedCounter показало свою практическую пользу при морфометрии зерен пшеницы и ячменя. Расширение функциональности за счет возможности оценки цветовых и текстурных характеристик оболочки зерна может стать существенным улучшением в практическом применении.

Методы и алгоритмы: В обновленной версии приложения SeedCounter была реализована поддержка различных протоколов съемки: с масштабированием на основе листа бумаги стандартного формата с коррекцией перспективы; на сплошном фоне с масштабированием на основе цветовой шкалы ColorChecker Mini Classic target (https://calibrite.com/us/product/colorchecker-classic-mini); на листе бумаги с цветовой шкалой.

Помимо этого, в новой версии приложения был реализован расчет цветовых и текстурных характеристик зерен [5]. Цветовые характеристики включают: средние значения интенсивности цветовых каналов, глобальную цветовую гистограмму, дескриптор доминантных цветов, дескриптор цветовой раскладки.

Текстурные характеристики включают: среднее, дисперсию, инерцию, гомогенность, единообразие, энергию, корреляцию, энтропию, максимальную вероятность, тень кластера, выпуклость кластера, рассчитываемые на основе матрицы совпадения уровней серого (GLCM, Grey Level Co-occurrence Matrix), а

также короткие серии, длинные серии, неравномерность уровня серого, неравномерность длин серий, процент серий, энтропию, рассчитываемые на основе матрицы длин серий уровней серого (GLRM, Gray Level Run-length Matrix). Как цветовые, так и текстурные характеристики вычисляются с учетом масок изображений зерен, что важно, так как при вычислении признаков на прямоугольной области пиксели фона могут вносить значительные искажения в рассчитываемые характеристики из-за малого размера зерен и их эллипсоидной формы.

Результаты: Обновленная версия приложения SeedCounter была использована для анализа краснозерных образцов озимой мягкой пшеницы. Проведенные оценки позволили сопоставить генетические вариации изученных образцов, цветовые характеристики зерен и их устойчивость к предуборочному прорастанию. На основе нашего метода было показано, что у краснозерных сортов существуют вариации по оттенкам цвета (голубизне и светлоте). В результате этой работы наравне с двенадцатью локусами, ассоциированными с признаками устойчивости к предуборочному прорастанию, были выявлены 26 локусов, ассоциированных с цветовыми характеристиками зерен (в том числе и новые).

Помимо этого, были проанализированы изображения зерен пшеницы популяции Международной инициативы по картированию Triticeae (International Triticeae Mapping Initiative, ITMI), включающие 116 генотипов (1740 изображений).

Анализ текстурных характеристик коллекции показал значимые различия отдельных генотипов (Synthetik_R93, Opata, ITMI_11, ITMI_103, ITMI_115, ITMI_42, ITMI_62, ITMI_82, ITMI_88) по таким характеристикам, как: среднее, гомогенность, единообразие, корреляция.

Выводы: Обновленное приложение SeedCounter расширяет возможности для определения цветовых и текстурных характеристик зерен злаковых. Цветовые признаки могут быть полезны для исследования содержания пигментов и метаболитов в оболочке зерна, таких как каротиноиды, антоцианы, меланины. Текстурные признаки могут быть полезны для оценки характеристик поверхности зерен, таких как шершавость.

Реализованные цветовые характеристики позволяют определять тонкие различия оттенков зерна даже между образцами, имеющими один тип окраски. Такой подход позволяет выявлять локусы, связанные с цветовыми признаками. Приложение SeedCounter позволило выявить новые локусы контроля цветовых характеристик зерен вне зависимости от признака устойчивости к предуборочному прорастанию.

Проведенный анализ текстурных признаков продемонстрировал, что в популяции мягкой пшеницы ITMI существует разнообразие зерен по степени шершавости оболочки зерен, которое можно выявлять и анализировать на основе предложенных алгоритмов оценки текстуры зерен.

Финансирование: Работа поддержана грантом Российского научного фонда (№ 22-74-00122) и Курчатовским геномным центром ИЦиГ СО РАН, Новосибирск, Россия.

SeedCounter app update – evaluation of color and texture characteristics of cereal grains

Komyshev E.G.[1]*, Genaev M.A.[1, 2], Afonnikov D.A.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *komyshev@bionet.nsc.ru*

Key words: mobile app; color characteristics; textural characteristics; cereal grains

Motivation and Aim: The grain coat is the main barrier between the grain and the external environment, so a number of important biological functions are associated with its characteristics: moisture absorption, grain viability, resistance to pre-harvest germination [1]. The color of cereal grain coat is an important trait, it characterizes the pigments and metabolites contained in it, such as carotenoids, anthocyanins, melanins.
In addition to color, the texture of the grain is also of interest. For example, the roughness of the grain coat in wheat is a trait that is related to the susceptibility of grains to pest damage [2], technological properties of grain [3], and moisture content [4]. When studying the mechanisms of genetic control of certain grain traits, breeders and geneticists face the need to evaluate the color and textural characteristics of their shells.
Mobile devices are increasingly being used in various fields of scientific observation, including agro-biotechnology. Thanks to the improvement of digital technology, the quality of color reproduction on modern camera models is becoming more and more accurate. This makes it possible to assess the color characteristics of plant organs with high quality. The previously developed SeedCounter application has shown its practical usefulness in morphometry of wheat and barley grains. Expanding the functionality to include the ability to assess the color and texture characteristics of the grain coat could be a significant improvement in practical applications.
Methods and Algorithms: In the updated version of the SeedCounter application, support for different imaging protocols was implemented: with scaling based on a standard-size sheet of paper with perspective correction; on a solid background with scaling based on the ColorChecker Mini Classic target (https://calibrite.com/us/product/colorchecker-classic-mini); on a sheet of paper with a color scale.
In addition, the calculation of color and texture characteristics of grains was implemented in the new version of the application [5]. Color characteristics include: average intensity values of color channels, global color histogram, descriptor of dominant colors, descriptor of color layout. Texture characteristics include: mean, variance, inertia, homogeneity, uniformity, energy, correlation, entropy, maximum probability, cluster shade, cluster prominance, calculated based on the Gray Level Co-occurrence Matrix (GLCM, Grey Level Co-occurrence Matrix), and short run, long run, gray level on-uniformity, run length non-uniformity, run ratio, entropy, calculated based on the Gray Level Run-length Matrix (GLRM, Gray Level Run-length Matrix).
Both color and texture features are computed by considering grain image masks, which is important because when computing features on a rectangular region, background pixels can introduce significant distortions in the computed features due to the small size of grains and their ellipsoidal shape.

Results: The updated version of the SeedCounter application was used to analyze red-grained winter soft wheat accessions. The evaluations made it possible to compare genetic variations of the studied accessions, color characteristics of grains and their resistance to preharvest germination. Based on our method, it was shown that variations in color shades (blueness and lightness) exist in red-grain varieties. As a result of this work, 26 loci associated with grain color traits (including new ones) were identified along with twelve loci associated with preharvest germination resistance traits.
In addition, wheat grain images of the International Triticeae Mapping Initiative (ITMI) population including 116 genotypes (1,740 images) were analyzed.
Analysis of the textural characteristics of the collection showed significant differences between individual genotypes (Synthetik_R93, Opata, ITMI_11, ITMI_103, ITMI_115, ITMI_42, ITMI_62, ITMI_82, ITMI_88) in terms of such characteristics as mean, homogeneity, uniformity, and correlation.
Conclusion: The updated SeedCounter application expands the possibilities for determining color and textural characteristics of cereal grains. Color traits can be useful to investigate the content of pigments and metabolites in the grain coat, such as carotenoids, anthocyanins, melanins. Textural attributes may be useful for evaluating grain surface characteristics such as roughness.
Realized color features allow the detection of subtle differences in grain hues even between samples having the same color type. This approach allows the identification of loci associated with color traits. The SeedCounter application allowed the identification of new loci controlling grain color traits independent of the preharvest germination resistance trait.
The analysis of textural traits demonstrated that in the ITMI soft wheat population there is grain diversity in terms of grain coat roughness, which can be detected and analyzed based on the proposed algorithms for grain texture evaluation.
Funding: The work was supported by RSF (project No. 22-74-00122) and Kurchatov Genomic Center, Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences.

Список литературы/References

1. Souza F.H., Marcos-Filho J. The seed coat as a modulator of seed-environment relationships in Fabaceae. *Braz J Botany*. 2001;24:365-375
2. D'Isita I. et al. Susceptibility of old and modern wheat genotypes to Sitophilus granarius (L.) and Rhyzoperta Dominica (F.). *J Stored Prod Res*. 2024;106:102265
3. Molenda M. et al. Friction of wheat grain. *Acta Agrophys*. 1995;(4):3-89
4. Królczyk J.B. Metrological changes in the surface morphology of cereal grains in the mixing process. *Int Agrophys*. 2016;30(2):193-202
5. Komyshev E.G., Genaev M.A., Afonnikov D.A. Analysis of color and texture characteristics of cereals on digital images. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2020;24(4):340-347

doi 10.18699/bgrs2024-06-13

Разработка и применение системы геномного редактирования ячменя (*Hordeum vulgare*) для получения трансгенных линий с целью изучения метаболизма жирных кислот

Короткова А.М.[1]*, Колосовская Е.В.[3], Вихорев А.В., Герасимова С.В., Хертиг К.[3], Кумлен Й.[3], Хлесткина Е.К.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Федеральный исследовательский центр «Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова» (ВИР), Санкт-Петербург, Россия*

[3] *Институт генетики растений и исследований сельскохозяйственных культур им. Лейбница, Гатерслебен, Германия*

* *korotkova@bionet.nsc.ru*

Ключевые слова: ячмень; CRISPR/Cas9; цитохром P450; кетоацил-КоА-синтетазы

Мотивация и цель: Ячмень (*Hordeum vulgare*) является одной из ключевых зерновых культур, широко используемых в сельском хозяйстве и пищевой промышленности. Анализ транскриптомных данных, полученных при изучении мутантов ячменя по гену *HvWin1*, кодирующему регуляторный белок семейства AP2/ERF и управляющему биосинтезом восков [1], позволил выделить четыре гена, проявляющих значительные изменения экспрессии в этих мутантах относительно растений дикого типа. Целью данного исследования является применение системы направленного геномного редактирования ячменя сорта Golden promise путем нокаута отобранных генов, непосредственно участвующих в метаболизме жирных кислот. К ним относятся ген, кодирующий цитохром P450, и три гена, кодирующих кетоацил-КоА-синтетазы.

Методы и алгоритмы: Для разработки векторов системы CRISPR/Cas9 использовались последовательности гена, кодирующего цитохром P450, и трех генов, кодирующих кетоацил-КоА-синтетазы, полученные из базы данных ячменя сорта Golden promise. Последовательности были проанализированы с помощью инструментов Benchling и WU-CRISPR для определения оптимальных участков для разрезания и разработки соответствующих направляющих РНК. Вторичная структура направляющих РНК была предсказана с использованием программы RNAfold. Сконструированные векторы экспрессии содержали кодирующую последовательность нуклеазы Cas9, оптимизированную для кукурузы, направляющую РНК и промотор U6 пшеницы *Triticum aestivum*. Финальные наборы были переклонированы в бинарный вектор p6i (DNA cloning service, Hamburg, Germany) и получили названия pEK7, pEK13, pEK17, pEK25.

Результаты: Полученные векторы экспрессии (pEK7, pEK13, pEK17, pEK25) были успешно использованы для трансформации незрелых зародышей ячменя сорта Golden promise методом агробактериальной трансформации. В результате было получено несколько трансгенных линий ячменя с мутациями в соответствующих генах: 23 растения-регенеранта для гена, кодирующего цитохром P450; а также 38, 44 и 23 растения-регенеранта для трех различных

генов, кодирующих кетоацил-КоА-синтетазы. Полученные трансгенные линии ячменя будут охарактеризованы в поколении Т1 для анализа изменений в составе жирных кислот по сравнению с контрольными растениями дикого типа.
Выводы: В результате данного исследования была успешно разработана и применена система CRISPR/Cas9 для редактирования генома ячменя сорта Golden promise с целью нокаута генов, непосредственно вовлеченных в метаболизм жирных кислот, включая ген, кодирующий цитохром Р450, и три гена, кодирующие кетоацил-КоА-синтетазы. Были получены многочисленные независимые трансгенные линии ячменя с мутациями в указанных генах, представляющие ценный материал для дальнейших исследований и манипуляций с биосинтезом жирных кислот в ячмене с целью улучшения качественных характеристик этой важной сельскохозяйственной культуры. Полученные результаты открывают перспективы для создания новых сортов ячменя с улучшенными характеристиками и могут значительно способствовать увеличению эффективности сельского хозяйства.
Финансирование: Работа выполнена при поддержке РНФ № 21-66-00012.

Development and application of genome editing system in barley (*Hordeum vulgare*) for generating transgenic lines to study fatty acid metabolism

Korotkova A.M.[1]*, Kolosovskaya E.V.[3], Vihorev A.V., Gerasimova S.V., Hertig C.W.[3], Kumlehn J.[3], Khlestkina E.K.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Federal Research Center N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia*

[3] *Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany*

* *korotkova@bionet.nsc.ru*

Key words: barley; CRISPR/Cas9; cytochrome P450; ketoacyl-CoA synthetases

Motivation and Aim: Barley (*Hordeum vulgare*) is one of the key cereal crops widely used in agriculture and the food industry. The analysis of transcriptomic data obtained from studying barley mutants for the *HvWin1* gene, encoding a regulatory protein of the AP2/ERF family and involved in wax biosynthesis [1], identified four genes that significantly changed their expression in these mutants. The aim of this study was to apply targeted genome editing in barley, specifically knockout of the selected genes directly involved in fatty acid metabolism. These genes include the cytochrome P450 gene and three genes encoding ketoacyl-CoA synthetases.
Methods and Algorithms: To develop CRISPR/Cas9 vector constructs, gene sequences encoding cytochrome P450 and three genes encoding ketoacyl-CoA synthetases were utilized. These sequences were obtained from the database of cultivar Golden promise. The sequences were analyzed using Benchling and WU-CRISPR tools to identify optimal cleavage sites and design corresponding guide RNAs (gRNAs). The secondary structure of the gRNAs was predicted using the RNAfold program. The constructed expression vectors contained the coding sequence of Cas9 nuclease optimized for maize, along with the gRNA and the U6 promoter from *Triticum aestivum*. The final constructs

were re-cloned into the binary vector p6i (DNA cloning service, Hamburg, Germany) and designated as pEK7, pEK13, pEK17, and pEK25.

Results: The obtained expression vectors (pEK7, pEK13, pEK17, pEK25) were successfully utilized for the transformation of immature embryos of Golden promise barley using the Agrobacterium-mediated transformation method. As a result, several transgenic barley lines carrying mutations in the respective genes were obtained: 23 regenerated plants for the cytochrome P450 gene; as well as 38, 44, and 23 regenerated plants for the three different genes encoding ketoacyl-CoA synthetases. These generated transgenic barley lines will be characterized in the T1 generation to analyze changes in fatty acid composition compared to wild-type control plants.

Conclusion: As a result of this study, a successful CRISPR/Cas9 system was developed and applied for genome editing in Golden promise barley, aiming to knock out genes directly involved in fatty acid metabolism, including the cytochrome P450 gene and three genes encoding ketoacyl-CoA synthetases. Numerous independent transgenic barley lines carrying mutations in these genes were generated, providing valuable material for further research and manipulation of fatty acid biosynthesis in barley to improve its qualitative traits. These findings open up avenues for developing new barley cultivars with enhanced characteristics, contributing significantly to agricultural efficiency.

Funding: The study is supported by the Russian Science Foundation (grant number 21-66-00012).

Список литературы/References

1. Gerasimova S.V., Kolosovskaya E.V., Vikhorev A.V. et al. WAX INDUCER 1 regulates β-diketone biosynthesis by mediating expression of the cer-cqu gene Cluster in barley. *Int J Mol Sci*. 2023;24:6762. doi 10.3390/ijms24076762

doi 10.18699/bgrs2024-06-14

Изменение экспрессии генов циркадного ритма в ответ на холодовой стресс у сортов томата *Solanum lycopersicum* L.

Кочиева Е.З.*, Щенникова А.В.

Институт биоинженерии ФИЦ «Фундаментальные основы биотехнологии Российской академии наук», Москва, Россия

** ekochieva@yandex.ru*

Ключевые слова: стрессовая память растения; циркадный ритм; Пасленовые

Мотивация и цель: В течение жизни растения подвергаются комплексному воздействию стрессовых факторов, что в случае сельскохозяйственных культур приводит к значительным потерям урожая. В ответ на стресс активируются защитные механизмы. При этом растения способны сохранять информацию о стрессовых событиях (явление наследуемой стрессовой памяти) и на основе полученного опыта оптимизировать свои реакции на повторные стрессы [1–3]. Целью нашей работы стало определение возможного участия генов циркадного ритма в ответе растений на стрессовые факторы у культурных видов сем. Пасленовые.

Методы и алгоритмы: Исследования проводили на сортах томата *Solanum lycopersicum* L., имеющих сходные морфологические характеристики и сроки созревания, но различающихся по холодостойкости. Проростки сортов, выращиваемые в нормальных условиях, подвергались имитации ночных похолоданий (краткосрочное воздействие низкой температуры (3 °C) в темновой фазе роста). В течение суточного цикла в проростке определялся профиль экспрессии (РВ-ПЦР) генов циркадного ритма. Далее растения подвергались повторным симуляциям холодового стресса; определялся профиль экспрессии генов, показавших дифференциальную экспрессию в первом цикле эксперимента. Контрольные растения параллельно выращивались в нормальных условиях. Одновременно с экспрессией генов в листьях проростков измерялось содержание хлорофиллов биохимическими методами. Данные экспрессионного и биохимического анализа обрабатывались статистически с учетом двух биологических и трех технических повторов.

Результаты: Адаптационная эволюция организмов, в том числе растений, связана с возникновением циркадных ритмов – повторяющихся колебаний в молекулярно-генетических процессах с периодом 24 ч, цикличностью «день–ночь» и сезонной спецификой [4]. Для оценки воздействия краткого похолодания в условиях ночи в листьях проростков трех сортов томата был определен уровень экспрессии генов циркадного ритма *GIGANTEA*, *FKF* и *CONSTANS*, а также генов фотосистем (*psaA*, *psbA*). В результате было обнаружено, что все анализируемые гены демонстрировали дифференциальную экспрессию в исследуемых образцах растений. Сравнение с контрольными растениями позволило оценить причины дифференциальной экспрессии, вычесть фотоэффект смены суточного ритма (день–ночь) и вычленить воздействие именно низкой температуры. Было показано, что низкая температура в разной степени усиливает циркадные колебания экспрессии каждого из целевых генов. При этом для содержания

хлорофиллов зависимость от температуры также присутствовала, но была менее выражена. В случае повторных воздействий холодового стресса в темновой фазе было обнаружено, что гены фотосистем характеризуются сходными между повторами стресса изменениями в экспрессионном уровне, тогда как гены циркадной системы меняли модель экспрессии. Используемые сорта томата, несмотря на схожесть морфологических характеристик, различались по силе ответа (экспрессия генов) на холодовой стресс и его повторы, что согласуется с предполагаемыми различиями степени стрессоустойчивости сортов.
Выводы: На основании полученных результатов можно предположить, что гены циркадного ритма (*GIGANTEA*, *FKF* и *CONSTANS*) и фотосистем (*psaA*, *psbA*) участвуют в реакции растений томата на холодовой стресс. Динамика экспрессии данных генов в ответ на похолодание может быть ассоциирована с холодостойкостью сортов томата.
Финансирование: Исследование поддержано грантом РНФ (№ 24-16-00043).

Changes in the expression of circadian rhythm genes in response to cold stress in tomato *Solanum lycopersicum* L. varieties

Kochieva E.Z.*, Shchennikova A.V.
Institute of Bioengineering, Federal Research Center "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, Russia
** ekochieva@yandex.ru*

Key words: stress memory; plants, circadian rhythm; Solanaceae

Motivation and Aim: During their lifetime, plants, including crops, are exposed to complex stress factors, which lead to significant yield losses and stimulate defense mechanisms. It is remarkable that plants are able to retain information about stressful events (the phenomenon of inherited stress memory) and, based on the experience gained, optimize their reactions to repeated stresses [1–3]. The aim of our work was to determine the possible involvement of circadian rhythm genes in the response of plants to stress factors in Solanaceae cultivated species.
Methods and Algorithms: Research was carried out on tomato varieties with similar morphological characteristics and maturation periods, but differing in resistance to cold stress. Seedlings grown under normal conditions were subjected to simulated nighttime cold stress (short-term exposure to low temperature (3 °C) in the dark growth phase). During the daily cycle, the expression profile of circadian rhythm genes was determined in the seedling by qRT-PCR. Next, the plants were subjected to repeated exposure of cold stress; the expression profile of genes showing differential expression in the first cycle of the experiment was determined. The control plants were grown in parallel under normal conditions Additional to gene expression, chlorophyll content in the leaves was measured using biochemical methods. All experiments were carried out in two biological and three technical repeats.
Results: The adaptive evolution of organisms, including plants, is associated with the circadian rhythms emergence – repetitive fluctuations in molecular genetic processes with a period of 24 hours, a day–night cycle and seasonal specificity [4]. To assess the effects of a short-time cold stress at night in seedlings of three tomato varieties, the expression level of the circadian rhythm genes *GIGANTEA*, *FKF* and *CONSTANS*, as

well as the genes of photosystems (*psaA*, *psbA*) was determined, and the chlorophyll (*a* and *b*) content was measured. As a result, it was found that all the analyzed genes showed differential expression in the studied tomato accessions. Comparison with control plants made it possible to assess the causes of differential expression, subtract the photoelectric effect of the circadian rhythm (day–night) change and reveal the effect of low temperature. It has been shown that exposure to low temperature increases circadian fluctuations in the expression of each of the target genes in various degrees. At the same time, temperature dependence was also present for the content of chlorophylls, but it was less pronounced. In the case of repeated cold stress effects in the dark phase, it was found that the photosystem genes are characterized by similar changes in the expression level in each cycle of stress repeats, whereas circadian system genes changed their expression pattern in relation to the level of change. The tomato varieties used, despite the similarity of morphological characteristics, differed in the strength of the response (gene expression) to cold stress and number of stress cycles, which is consistent with the assumed differences in the degree of stress resistance of these varieties.

Conclusion: Based on the results obtained, it can be assumed that the genes of circadian rhythm (*GIGANTEA*, *FKF* and *CONSTANS*) and photosystems (*psaA*, *psbA*) are involved in the response of tomato plants to cold stress. The dynamics of expression of these genes in cold stress response may be associated with the cold resistance of tomato varieties.

Funding: The study was supported by Russian Science Foundation (grant No. 24-16-00043).

Список литературы/References

1. Li Y., Luo Z.Q., Yuan J., Wang S., Liu J., Su P., Zhou J.H., Li X., Yang J., Guo L.P. Metabolic and transcriptional stress memory in sorbus pohuashanensis suspension cells induced by yeast extract. *Cells.* 2022;11(23):3757. doi 10.3390/cells11233757
2. Nishad A., Nandi A.K. Recent advances in plant thermomemory. *Plant Cell Rep.* 2021;40(1):19-27. doi 10.1007/s00299-020-02604-1
3. Zuo D.D., Ahammed G.J., Guo D.L. Plant transcriptional memory and associated mechanism of abiotic stress tolerance. *Plant Physiol Biochem.* 2023;201:107917. doi 10.1016/j.plaphy.2023.107917
4. Hut R.A., Beersma D.G. Evolution of time-keeping mechanisms: early emergence and adaptation to photoperiod. *Philos Trans R Soc Lond B Biol. Sci*. 2011;366:2141. doi 10.1098/rstb.2010.0409

doi 10.18699/bgrs2024-06-15

Предсказание функции белков на основе гомологии

Малюгин Е.[1, 3]*, Мустафин З.[2, 3], Пронозин А.[2, 3], Генаев М.[2, 3], Афонников Д.[2, 3]

[1] *Новосибирский государственный университет, Новосибирск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[3] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

* *evgeny.malyugin98@gmail.com*

Ключевые слова: предсказание функции; генная онтология; возраст генов; гомология; ортогруппы

Мотивация и цель: С появлением методов секвенирования нового поколения число секвенированных последовательностей геномов растет огромными темпами. Источниками поступления такого большого количества данных являются различные геномные, метагеномные и транскриптомные проекты. Все это позволяет распознавать структуру генома и предсказывать локализацию белок-кодирующих генов с высокой точностью. Однако для распознанных генов функция, как правило, оказывается неизвестной. Существующие для решения этой задачи методы преимущественно используют для поиска гомологичных последовательностей с известными функциями, описанными в терминах Gene Ontology. Возраст генов имеет прямое отношение к задаче аннотации функций генов: в задачах предсказания функции белков по гомологии важной проблемой является наличие аннотированных гомологов для искомых последовательностей в базах данных. Очевидно, что если гомологичные последовательности для исходной последовательности отсутствуют, то это не позволяет получить информацию для ее аннотации [1, 2]. Такая ситуация характерна для молодых или таксон-специфичных (орфанных) генов, представленных в одном или нескольких ближайших видах и отсутствующих у других организмов. Для них предсказание функции затруднено, и его точность оказывается ниже, чем для древних генов, гомологи для которых представлены более широко. Таким образом, создание метода аннотации функций генов, который бы работал с высокой точностью вне зависимости от возраста генов, является достаточно актуальным.

Методы и алгоритмы: Мы предложили алгоритм аннотации функций генов на основе поиска сходных последовательностей методом *k* ближайших гомологов, анализа ортологических групп в базе данных OrthoDB v. 10.1 [3] и фильтрации терминов на основе сходства последовательностей методом логистической регрессии. Поиск гомологичных последовательностей проводился в базе данных с использованием алгоритма USEARCH v 11.0.667 [4]. Методы были опробованы на последовательностях белок-кодирующих генов ряда геномов растений, включая *Arabidopsis thaliana* и др. Оценка возрастов генов была использована с помощью программы Orthoscape [5]. Мы также предложили использовать для оценки точности методов предсказания функций генов меру эффективности предсказания (PE), которая позволяет избежать влияния использования различных аннотаций при сравнении разных методов предсказания.

Результаты: Точность предложенного алгоритма при анализе генома *A. thaliana*, *Oryza sativa* и *Zea mays* оказалась сравнимой с методами Blast2GO [6] и eggNOG [7]. Так, эффективность предсказания (PE) на данных *A. thaliana* для нашего

метода составила 85.03, 87.17, 89.24 % для генов молодого, среднего и старого возраста соответственно против 85.32, 84.48, 81.69 % для метода Blast2GO. Показано, что при определенных параметрах алгоритма точность предсказания функций у генов разных возрастов (в том числе молодых) оказывается близкой. Таким образом, предложенный метод позволяет нивелировать различия в точности предсказания функции для генов разных возрастов, при этом в лучшую сторону.

Выводы: Точность предсказания генов молодого возраста гораздо ниже, чем точность предсказания средних и старых генов. Использование ортогрупп и метода логистической регрессии позволяет повысить точность предсказания для генов молодого возраста.

Финансирование: Исследование поддержано Курчатовским геномным центром Института цитологии и генетики СО РАН, соглашение с Министерством образования и науки Российской Федерации № 075-15-2019-1662.

Predicting protein function based on homology

Malyugin E.[1, 3]*, Mustafin Z.[2, 3], Pronozin A.[2, 3], Genaev M.[2, 3], Afonnikov D.[2, 3]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

** evgeny.malyugin98@gmail.com*

Key words: function prediction; gene onthology; gene age; homology; orthogroups

Motivation and Aim: With the advent of next-generation sequencing methods, the number of sequenced genome sequences is growing at a tremendous rate. Various genomic, metagenomic, and transcriptomic projects are the sources of this large amount of data. All of these make it possible to recognise genome structure and predict the localisation of protein-coding genes with high accuracy. However, for novel genes, the function is usually unknown. The methods to predict function for novel genes predominantly use the search for homologous sequences with known functions described in terms of Gene Ontology. The age of genes is directly related to the task of gene function annotation: in the tasks of protein function prediction by homology, an important problem is the availability of annotated homologues for the searched sequences in databases. Obviously, in the case if homologous sequences for the query sequence are not available, gene function prediction for these methods is difficult [1, 2]. This situation is typical for young or taxon-specific (orphan) genes represented in one or more relative species and absent in other organisms. For them, function prediction is difficult and its accuracy is lower than for ancient genes for which homologues are more widely represented. Thus, the creation of a method of gene function annotation that would work with high accuracy regardless of the age of genes is quite relevant.

Methods and Algorithms: We proposed an algorithm for gene function annotation based on searching similar sequences by *k* nearest homologues method, analysing orthologous groups in OrthoDB database v. 10.1 [3] and filtering terms based on sequence similarity by logistic regression. Homologous sequences were searched in the database using the USEARCH v 11.0.667 algorithm [4]. The methods were tested on sequences of protein-coding genes from a number of plant genomes including *Arabidopsis thaliana* and

several other plant species. Estimation of gene ages was performed by the Orthoscape method [5]. We also proposed a novel prediction performance measure, the prediction efficiency which is independent on the usage of different annotation source for different prediction methods.

Results: The accuracy of the proposed algorithm in analysing the genome of *A. thaliana*, *Oryza sativa* and *Zea mays* was comparable to the Blast2GO and eggNOG methods [6, 7]. Thus, the prediction efficiency (PE) on *A. thaliana* data for our method was 85.03 %, 87.17 %, 89.24 % for young, middle-aged and old genes respectively, compared to 85.32 %, 84.48 %, 81.69 % for the Blast2GO method. It is shown that at certain parameters of the algorithm the accuracy of function prediction for genes of different ages (including young genes) is close. Thus, the proposed method allows levelling the differences in the accuracy of function prediction for genes of different ages, at the same time for the better.

Conclusion: The prediction accuracy of young age genes is much lower than the prediction accuracy of middle-aged and old genes. The use of orthogroups and logistic regression method can improve the prediction accuracy for young age genes.

Funding: The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation, No. 075-15-2019-1662.

Список литературы/References

1. Liebeskind B.J., McWhite Claire D., Marcotte E.M. Towards consensus gene ages. *Genome Biol Evol*. 2016;8(6):1812-1823
2. Mustafin Z.S., Zamyatin V.I., Konstantinov D.K., Doroshkov A.V., Lashin S.A., Afonnikov D.A. Phylostratigraphic analysis shows the earliest origination of the abiotic stress associated genes in *A. thaliana*. *Genes (Basel)*. 2019;10(12):963. doi 10.3390/genes10120963
3. Kriventseva E.V., Kuznetsov D., Tegenfeldt F., Manni M., Dias R., Simão F.A., Zdobnov E.M. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res*. 2019;10(12):963. doi 10.3390/genes10120963
4. Robert C.E. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460-2461
5. Mustafin Z.S., Lashin S.A., Matushkin Yu.G., Gunbin K.V., Afonnikov D.A. Orthoscape: a Cytoscape plugin for grouping and visualization KEGG based gene networks by taxonomy and homology principles. *BMC Bioinformatics*. 2017;18(Suppl. 1):1427
6. Conesa A., Götz S., García-Gómez J.M., Terol J., Talón M., Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21(18):3674-3676
7. Huerta-Cepas J., Szklarczyk D., Heller D. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47(1):309-314

doi 10.18699/bgrs2024-06-16

Различные структурные варианты сайта связывания транскрипционного фактора EIN3 опосредуют регуляцию разных биологических процессов в ответе на этилен у *Arabidopsis thaliana* L.

Маслакова А.А.[1], Долгих В.А.[1, 2], Землянская Е.В.[1, 2] *

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *ezemlyanskaya@bionet.nsc.ru*

Ключевые слова: транскрипционный фактор ETHYLENE-INSENSITIVE3 (EIN3); сайты связывания EIN3 (EIN3 binding site; EBS); генная онтология

Мотивация и цель: Транскрипционный фактор ETHYLENE-INSENSITIVE3 (EIN3) играет ключевую роль в реакции растений на фитогормон этилен. Под действием этилена он связывается с определенными участками ДНК в промоторах генов, называемыми EBS (EIN3 binding site), запуская таким образом транскрипционный ответ генов [1]. Ранее нами было показано, что существуют четыре структурных варианта EBS, два из них представляют собой инвертированный повтор классического EBS с перекрыванием копий на 1 и 2 нуклеотида (2EBS[-1] и 2EBS[-2] соответственно) [2]. Предсказано, что свойства этих мотивов с точки зрения регуляции экспрессии генов в ответ на этилен различаются. В данной работе мы исследовали, связаны ли различные структурные варианты EBS с регуляцией специфических признаков под действием этилена.

Методы и алгоритмы: Для определения потенциальных цис-элементов двух типов в промоторах этилен-чувствительных генов мы использовали публично доступные данные экспериментов ChIP-seq для EIN3 и RNA-seq по обработке 3-дневных этиолированных проростков *Arabidopsis thaliana* этиленом в течение 1, 4 и 12 часов [3]. Отбор пиков ТФ EIN3, расположенных в регуляторных районах генов [–1500; +1), проводили с помощью сервиса PAVIS [4]. Распознавание мотивов 2EBS[-1] и 2EBS[-2] в пиках проводили с помощью конвейера SFmotif [5]. Функциональная аннотация списков этилен-чувствительных генов-мишеней EIN3 с различными структурными вариантами EBS в промоторе проводилась с помощью онлайн-сервисов DAVID [6, 7] и AgriGO (v2.0) [8]. Кластеризация и визуализация терминов ГО проводились с помощью онлайн-инструмента REVIGO (v 1.8.1) [9].

Результаты: Было распознано 63 мотива 2EBS[-1] и 33 мотива 2EBS[-2] в регуляторных районах 36 и 25 активируемых этиленом генов-мишеней EIN3 соответственно. Функциональная аннотация списков генов с помощью сервисов AgriGO и DAVID выявила 45 и 158 обогащенных ГО-терминов в списках генов-мишеней с мотивом 2EBS[-1] и с мотивом 2EBS[-2] соответственно. Выявлено 33 общих ГО-термина, которые связаны с ответом на эндогенные стимулы (в том числе ответ на гормоны), регуляцией клеточных процессов (включая клеточные взаимодействия) и передачей сигнала (в том числе сигнальный путь,

активируемый этиленом, и передача сигнала фосфорилирования). Уникальными биологическими процессами для активируемых этиленом генов-мишеней EIN3 с мотивом 2EBS[-1] в промоторе были ответ на биотические стимулы, неорганические вещества, позитивная регуляция процессов развития многоклеточных организмов и регуляция цитокинин-активируемого сигнального пути. Уникальными биологическими процессами для активируемых этиленом генов-мишеней EIN3 с мотивом 2EBS[-2] в промоторе были процессы, связанные с регуляцией метаболизма и биосинтезом, с регуляцией экспрессии генов и процессинга РНК, ответом на абиотические стимулы (в том числе на этилен и снижение уровня кислорода), иммунный ответ на внешние воздействия, с развитием анатомических структур растений (включая постэмбриональное развитие) и процессами размножения и старения (в том числе листьев).
Выводы: Полученные результаты свидетельствуют о том, что различные структурные варианты EBS могут быть связаны с регуляцией специфических признаков под действием этилена.
Финансирование: Исследование поддержано грантом РНФ № 20-14-00140.

Structural variants of EIN3 binding site mediate regulation of distinct biological processes in response to ethylene in *Arabidopsis thaliana* L.

Maslakova A.A.[1], Dolgikh V.A.[1, 2], Zemlyanskaya E.V.[1, 2]*

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *ezemlyanskaya@bionet.nsc.ru*

Key words: ETHYLENE-INSENSITIVE3 (EIN3) transcription factor; EIN3 binding sites; gene ontology

Motivation and Aim: ETHYLENE-INSENSITIVE3 (EIN3) transcription factor plays a pivotal role in the plant response to phytohormone ethylene. Upon ethylene exposure, EIN3 binds to specific regions of DNA in gene promoters designated as EBS (EIN3 binding site), thereby initiating the transcriptional response of genes [1]. Our previous studies have demonstrated the existence of four distinct structural variants of EBS. Two of these variants are inverted repeats of the classic EBS motif with the overlap of 1 and 2 nucleotides between the copies (2EBS [-1] and 2EBS [-2], respectively). It was predicted that the regulatory properties of these motifs in terms of gene expression in response to ethylene differ. The objective of this study was to determine whether various structural EBS variants are associated with the regulation of specific traits in response to ethylene.
Methods and Algorithms: To identify the potential *cis*-elements of each type in the promoters of ethylene-sensitive genes, we used publicly available EIN3 ChIP-seq data and RNA-seq data on the transcriptomes of 3-day-old etiolated *Arabidopsis thaliana* seedlings treated with ethylene for 1, 4, and 12 hours [3]. We identified EIN3 peaks, which fall into the regulatory gene regions [–1500; +1) using the PAVIS service [4]. The recognition of 2EBS[-1] and 2EBS[-2] motifs in the peaks was conducted using SFmotif pipeline [5]. The functional annotation of lists of ethylene-sensitive EIN3 target genes, which possess different structural variants of EBS in their promoters, was performed

using the online services DAVID [6, 7] and AgriGO (v2.0) [8]. The clustering and visualization of Gene Ontology (GO) terms was conducted using the online tool REVIGO (v 1.8.1) [9].

Results: A total of 63 2EBS[-1] and 33 2EBS[-2] motifs were identified in the regulatory regions of 36 and 25 ethylene-activated EIN3 target genes, respectively. The functional annotation of gene lists using the AgriGO and DAVID services revealed 45 and 158 enriched GO terms in the lists of target genes with the 2EBS[-1] motif and with the 2EBS[-2] motif, respectively. A total of 33 common GO terms were identified, which were related to the response to endogenous stimuli (including the response to hormones), regulation of cellular processes (including cellular interactions) and signal transmission (including the ethylene-activated signaling pathway and phosphorylation signal transmission). The unique biological processes for ethylene-activated EIN3 target genes with the 2EBS[-1] motif in the promoter were response to biotic stimuli, inorganic substances, positive regulation of the development of multicellular organisms, and regulation of the cytokinin-activated signaling pathway. The unique biological processes for ethylene-activated EIN3 target genes with the 2EBS[-2] motif in the promoter were those related to the regulation of metabolism and biosynthesis, regulation of gene expression and RNA processing, response to abiotic stimuli (including ethylene and reduction of oxygen levels), immune response to external influences and the development of anatomical features. These processes also included development of plant anatomic structures (including postembryonic development) and reproduction and aging processes (including leaves).

Conclusion: The results obtained indicate that various structural variants of EBS can be associated with the regulation of specific traits under the action of ethylene.

Funding: The study was supported by the RSF grant No. 20-14-00140.

Список литературы/References

1. Solano R., Stepanova A., Chao Q., Ecker J.R. Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes Dev.* 1998;12(23):3703-3714. doi 10.1101/gad.12.23.3703
2. Dolgikh V., Levitsky V., Zemlyanskaya E., Oshchepkov D. EIN3 binding site architecture guides transcriptional response to ethylene in Arabidopsis. In: Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2020). Novosibirsk, 2020;305
3. Chang K.N., Zhong S., Weirauch M.T. et al. Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. *Elife*. 2013;2:e00675. doi 10.7554/eLife.00675
4. Huang W., Loganantharaj R., Schroeder B., Fargo D., Li L. PAVIS: a tool for peak annotation and visualization. *Bioinformatics*. 2013;29(23):3097-3099. doi 10.1093/bioinformatics/btt520
5. Долгих В.А., Землянская Е.В. Программный комплекс для анализа сайтов связывания транскрипционных факторов (АССТФ/SFmotif). Pat. 2023660670. ФГБНУ ФИЦ ИЦиГ СО РАН, 2023
6. Sherman B.T., Hao M., Qiu J., Jiao X., Baseler M.W., Lane H.C., Imamichi T., Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res*. 2022;50(W1):W216-W221. doi 10.1093/nar/gkac194
7. Huang da W., Sherman B.T., Lempicki R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57. doi 10.1038/nprot.2008.211
8. Tian T., Liu Y., Yan H., You Q., Yi X., Du Z., Xu W., Su Z. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res*. 2017;45(W1):W122-W129. doi 10.1093/nar/gkx382
9. Supek F., Bošnjak M., Škunca N., Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;6(7):e21800. doi 10.1371/journal.pone.0021800

doi 10.18699/bgrs2024-6-16

Метилирование ДНК – возможность адаптации к абиотическим воздействиям среды

Минасбекян Л.А.*, Авагян И.А.
Научно-исследовательский институт биологии, Ереванский государственный университет, Ереван, Армения
**minlia@ysu.am*

Key words: мягкая пшеница; ДНК метилирование; адаптация; абиотический фактор

Мотивация и цель: Целью статьи было показать воздействие одного из абиотических факторов среды, который все больше распространяется благодаря созданию все более плотной сети беспроводной коммуникационной связи, что в свою очередь воздействует на растения и их урожайность. В данной статье нами показано, что воздействия мм-волн на ранних стадиях развития проростков оставляют свой след на геноме растений, и основным маркером является в первую очередь ДНК метилирование.
Методы и алгоритмы: Для получения образцов, облученных мм-волнами, через сутки замоченные семена подвергались обработке мм-волнами различной экспозиции в интервале частот 45–53 ГГц в течение 20 мин генератором высоких частот Г4-141 (производство России) при плотности излучения 0.64 мВ/см^2, на расстоянии не более 18 см от излучателя, что обеспечивает покрытие всей площади чашек Петри с замоченными в один слой семенами. Обработанные различными частотами семена высаживались на отдельные лотки и проращивали далее при 270 °С в термостате еще 72 часа. Часть замоченных семян после каждой отдельной обработки различными частотами ЭМИ была высажена, и получен урожай пшеницы в последующей генерации. На схеме графически изображена последовательность действий в наших исследованиях.
Результаты: В естественных условиях произрастания на рост растения воздействуют ряд биотических и абиотических факторов. Эпигенетические изменения в геноме растений включают в себя различного рода обратимые химические модификации, происходящие как на самой ДНК, так и со взаимодействующими с нею белками. В результате происходят конформационные изменения хроматина, без изменения последовательности нуклеиновых остатков ДНК.
На проростках пшеницы *Triticum aestivum* сорта Воскеаск показана значимость роли метилирования ДНК в формировании ответа на абиотический стресс, а также наследование этих изменений в следующем поколении. Исследования проводились при воздействии крайне высоких частот электромагнитного излучения в интервале частот 45–53 ГГц. Изменения метилирования под воздействием стрессового абиотического фактора наблюдались в сторону как возрастания, так и понижения, что позволяет говорить об адаптивном значении этих изменений для мобилизации имеющихся ресурсов растения. При воздействии КВЧ ЭМИ в диапазоне частот 45–46 ГГц было получено повышение содержания 5мЦ в ДНК проростков, что объясняется ускорением метаболических процессов, роста и дифференциации тканей, вызванным воздействием стрессорного фактора. Однако более высокие частоты КВЧ ЭМИ в диапазоне 49–50.3 ГГц приводили к понижению

ДНК метилирования, что свидетельствовало о сильном стрессовом воздействии мм-волн этого диапазона, приводящем к нарушению поддерживающего метилирования в генах, репрессированных во время дифференцировки на ранних стадиях прорастания семени. Метилирование ДНК предоставляет селективные преимущества фенотипической пластичности и трансгенерационных эффектов, различающие разные виды растений в зависимости от величины и плоидности генома.

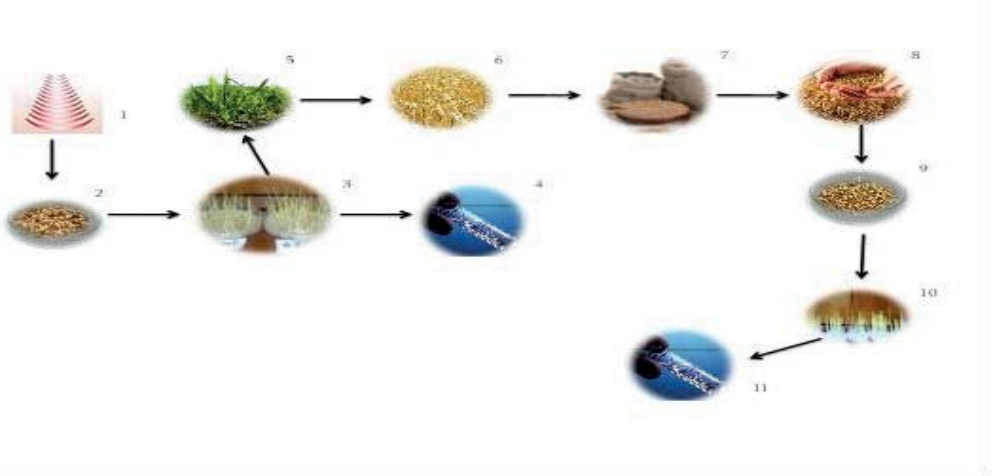


Рис. 1. Схема и последовательность исследований ДНК метилирования в поколениях I и II проростков пшеницы. 1 – источник облучения, 2 – замоченные на ночь семена, 3 – получение 4-суточных проростков (поколение I), облученных частотами в интервале 45–53 ГГц, 4 – выделение ДНК и определение степени метилирования, 5 – высадка облученных проростков в землю, 6–9 – получение урожая из проростков (поколение II), облученных различными экспозициями, 10 – получение проростков из урожая вышеуказанных проростков без какой-либо обработки, 11 – выделение ДНК и определение метилирования

Выводы: Эпигенетическая регуляция включает метилирование ДНК, модификации гистонов и ремоделирование хроматина, которые влияют на многие важные функции клетки, включая регуляцию экспрессии генов и поддержание целостности генома. Таким образом, нами показано, что воздействия мм-волн на ранних стадиях развития проростков оставляют свой «след» на геноме растений, основным маркером которого является ДНК метилирование. Нами также выявлена наследуемость эпигенетических изменений в растениях, выращенных из уже облученных проростков пшеницы, у которых сохранилась степень изменений метилирования проростков семян (второе поколение). Таким образом, показана важная роль ДНК метилирования как достоверного чувствительного биомаркера воздействия факторов среды на биологические организмы и его наследуемость из поколения в поколение, что позволяет сделать вывод об адаптивном значении этих изменений для мобилизации имеющихся ресурсов растения.

Title DNA methylation – the possibility of adaptation to abiotic environmental influences

Minasbekyan L.A.*, Avagyan I.A.

Research Institute of Biology, Yerevan State University, Yerevan, Armenia

** minlia@ysu.am*

Key words: bread wheat; DNA methylation; adaptation; abiotic factor

Motivation and Aim: The purpose of this article was to show the impact of one of the abiotic environmental factors that is increasingly spreading due to the creation of increasingly dense wireless communication networks, which in turn affects plants and their productivity. In this article, we show that exposure to mm waves at the early stages of seedling development leaves its mark on the plant genome, and the main marker is, first of all, DNA methylation.
Methods and Algorithms: To obtain samples irradiated with mm waves after 24 hours, soaked seeds were treated with mm waves of various exposures in the frequency range 45–53 GHz for 20 minutes using a G4-141 high-frequency generator (made in Russia) at a radiation density of 0.64 mV/cm^2, at a distance of no more than 18 cm from the emitter, which ensures coverage of the entire area of the petri dishes with seeds soaked in one layer. Seeds treated with different frequencies were planted on separate trays and further germinated at 270 °C in a thermostat for another 72 hours. Part of the soaked seeds, after each individual treatment with different frequencies of EMR, was planted and a wheat harvest was obtained in the subsequent generation. The diagram graphically depicts the sequence of actions in our research.

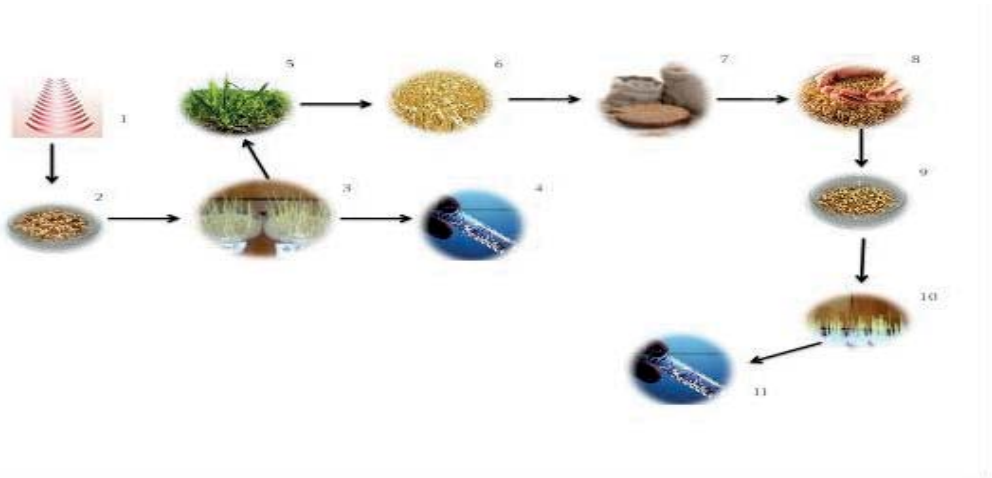


Fig. 1. A diagram and sequence of studies of DNA methylation in the I and II generations of wheat seedlings, where: 1 – irradiation source, 2 – seeds soaked overnight, 3 – obtaining 4-day-old seedlings (I generation), irradiated with frequencies in the range 45–53 GHz, 4 – DNA isolation and determination of the degree of methylation, 5 – planting irradiated seedlings in the ground, 6–9 – obtaining a harvest from seedlings (II generation) irradiated with various irradiations, 10 – obtaining seedlings from the harvest of the above seedlings without any processing and, finally, 11 – DNA extraction and methylation determination

Results: Under natural growing conditions, plant growth is affected by a number of biotic and abiotic factors. Epigenetic changes in the plant genome include various types of reversible chemical modifications that occur both on the DNA itself and on proteins

interacting with it. As a result, conformational changes in chromatin occur without changing the sequence of DNA nucleic acid residues.

Using wheat seedlings *Triticum aestivum*, variety Voskeask, the significance of the role of DNA methylation in the formation of a response to abiotic stress, as well as the subsequent inheritance of these changes in the next generation, was shown. The studies were carried out under exposure to extremely high frequencies of electromagnetic radiation in the frequency range 45–53 GHz. Changes in methylation under the influence of an abiotic stress factor changed both upward and downward, which allows us to conclude the adaptive significance of these changes for the mobilization of available plant resources. When exposed to EHF EMR in the frequency range 45–46 GHz, an increase in the content of 5mC in the DNA of seedlings was obtained, which is explained by the acceleration of metabolic processes, growth and differentiation of tissues caused by exposure to a stress factor. However, higher frequencies of EHF EMR in the range of 49–50.3 GHz led to a decrease in DNA methylation, which indicated a strong stressful effect of mm waves in this range, leading to disruption of maintenance methylation in genes repressed during differentiation at the early stages of seed germination. DNA methylation provides selective advantages for phenotypic plasticity and transgenerational effects, distinguishing different plant species depending on the size and ploidy of the genome.

Conclusion: Epigenetic regulation involves DNA methylation, histone modifications, and chromatin remodeling, which influence many important cellular functions, including gene expression regulation and genome integrity maintenance. Thus, we have shown that the effects of mm waves at the early stages of seedling development leave their "mark" on the plant genome, the main marker of which is DNA methylation. We also identified the heritability of epigenetic changes in plants grown from already irradiated wheat seedlings and the degree of changes in the methylation of seed seedlings (second generation), which was preserved. Thus, the important role of DNA methylation is shown as a reliable, sensitive biomarker of the impact of environmental factors on biological organisms and its heritability from generation to generation, which allows us to conclude the adaptive significance of these changes for the mobilization of available plant resources.

Список литературы/References

1. Bernstein C. DNA methylation and establishing memory. *Epigenet Insights*. 2022;19(15): 25168657211072499. doi 10.1177/25168657211072499
2. Pall M.L. Electromagnetic fields act similarity in plants as in animals: probable activation of calcium channels via their. *Current Chem Biol*. 2016;10(1):74-82. doi 10.2174/2212796810666160419160433
3. Springer N.M., Lisch D., Li Q. Creating order from chaos: epigenome dynamics in plants with complex genomes. *Plant Cell*. 2016;28(2):314-325
4. Vian A., Davies E., Gendraud M., Bonnet P. Plant responses to high frequency electromagnetic fields *Biomed Res Int*. 2016;2016:1830262. doi 10.1155/2016/1830262
5. Yuan G.C. Linking genome to epigenome. *Wiley Interdiscip Rev Syst Biol Med*. 2012;4(3):297-309. doi 10.1002/wsbm.1165

doi 10.18699/bgrs2024-06-17

Разработка биоинформатического конвейера для обработки данных, полученных методом GBS

Пронозин А.[1, 2]*, Афонников Д.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

** pronozinartem95@gmail.com*

Ключевые слова: GBS; ячмень; биоинформатический конвейер

Мотивация и цели: Развитие технологий секвенирования нового поколения открыло новые возможности для генотипирования различных организмов, включая растения. Метод генотипирования путем секвенирования (GBS) применяется для идентификации генетической изменчивости и более быстрого генотипирования образцов, а также является более экономически эффективным методом в сравнении с полногеномным секвенированием. GBS продемонстрировал свою надежность и гибкость для ряда видов и популяций растений. Этот метод был применен для генетического картирования, выявления молекулярных маркеров, геномной селекции, в исследовании генетического разнообразия, идентификации сортов, а также в исследованиях в области биологии-охраны природы и эволюционной экологии.

GBS существенно сокращает как стоимость, так и время, необходимое для секвенирования исследуемых образцов. Это привело к потребности в разработке качественного биоинформатического анализа для постоянно расширяющегося количества секвенированных данных.

В настоящей работе нами был разработан биоинформатический конвейер GBS-DP для анализа данных, полученных методом GBS. Конвейер применим для любых видов организмов. Конвейер полностью автоматизирован и позволяет обрабатывать большие объемы данных (более 400 образцов).

Методы и алгоритмы: Конвейер состоит из трех основных этапов: предобработка данных, поиск полиморфизмов, анализ генетического разнообразия. Реализация конвейера на платформе Snakemake позволила полностью автоматизировать процесс расчета и установки необходимых программных пакетов. Для тестового применения конвейера GBS-DP в настоящей работе был использован проект PRJEB39633 из базы данных European Nucleotide Archive (ENA) [1], который содержит библиотеки GBS для популяции ячменя.

Результаты: Конвейер предоставляет результаты оценки базовых характеристик секвенированных библиотек: длина прочтения для каждой библиотеки, средняя глубина прочтения, количество прочтений, приходящихся на одну библиотеку, среднее покрытие каждой библиотеки и всех библиотек в целом.

Также конвейер предоставляет результаты поиска полиморфизмов между исследуемыми генотипами. Для 272 исследуемых образцов выявлено 447409 SNP. Общее количество индел 46557. Параметр неравновесия по сцеплению (LD) (r2) был выбран равным 0.5. После применения фильтра LD осталось 45402 полиморфных и независимых SNP.

Конвейер предоставляет распределение выявленных SNP по хромосомам, анализ главных компонент генотипов на основе выявленных SNP и построение филогенетического дерева. Результаты анализа главных компонент на основе 45402 SNP показывают, что внутри исследованной популяции на диаграмме рассеяния в пространстве двух первых компонент четко выделяется несколько кластеров.
Выводы: В настоящей работе нами был предложен биоинформатический конвейер GBS-DP, который позволяет обрабатывать данные широкомасштабного секвенирования, проведенного методом GBS. Результаты демонстрируют достаточно высокую скорость работы конвейера как для больших данных (более 400 библиотек), так и для малых (30 библиотек). Конвейер предоставляет также анализ выявленных полиморфизмов.
Финансирование: Работа выполнена за счет финансирования Курчатовского геномного центра ФИЦ ИЦиГ СО РАН, соглашение с Министерством образования и науки РФ № 075-15-2019-1662. Вычисления проводились с использованием ресурсов ЦКП «Биоинформатика».

Development of a bioinformatics pipeline for GBS data processing

Pronozin A.[1, 2]*, Afonnikov D.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *pronozinartem95@gmail.com*

Key words: GBS; barley; bioinformatics pipeline

Motivation and Aim: The development of next-generation sequencing technologies has opened up new opportunities for genotyping various organisms, including plants. Genotyping by sequencing (GBS) is used to identify genetic variability and to genotype samples more rapidly, and is more cost-effective than whole-genome sequencing. GBS has demonstrated its reliability and flexibility for a number of plant species and populations. The method has been applied to genetic mapping, molecular marker detection, genomic selection, in genetic diversity studies, variety identification, and in conservation biology and evolutionary ecology studies.
GBS significantly reduces both the cost and the time required for sequencing the samples under study. This has led to the need to develop high-quality bioinformatics analysis for the ever-expanding amount of sequenced data.
In the present work, we have developed the GBS-DP bioinformatics pipeline for analyzing GBS-derived data. The pipeline is applicable to any species of organisms. The pipeline is fully automated and allows processing large amounts of data (more than 400 samples).
Methods and Algorithms: The pipeline consists of three main stages: data preprocessing, search for polymorphisms, and genetic diversity analysis. The implementation of the pipeline on the Snakemake platform allowed us to fully automate the process of calculation and installation of the necessary software packages. For the test application of the GBS-DP pipeline in this work, we used the PRJEB39633 project from the

European Nucleotide Archive (ENA) database [1], which contains GBS libraries for the barley population.

Results: The pipeline provides results of evaluation of basic characteristics of sequenced libraries: read length for each library, average read depth, number of reads per library, average coverage of each library and all libraries in general.

The pipeline also provides the results of the search for polymorphisms between the genotypes analyzed. For the 272 samples analyzed, 447,409 SNPs were identified. The total number of indels is 46,557. The linkage disequilibrium (LD) parameter (r2) was chosen to be 0.5. After applying the LD filter, 45,402 polymorphic and independent SNPs remained.

The pipeline provides the distribution of the identified SNPs across chromosomes, principal component analysis of genotypes based on the identified SNPs, and construction of a phylogenetic tree. The results of principal component analysis based on 45,402 SNPs show that within the studied population, several clusters are clearly distinguished in the scatter diagram in the space of the first two components.

Conclusion: In the present work, we proposed the GBS-DP bioinformatics pipeline, which allows us to process large-scale sequencing data performed by the GBS method. The results demonstrate quite high speed of the pipeline for both large data (more than 400 libraries) and small data (30 libraries). The pipeline also provides analysis of detected polymorphisms.

Funding: This work was funded by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation No. 075-15-2019-1662. Calculations were performed using the resources of the Bioinformatics Center.

Список литературы/References

1. Leinonen R. et al. The European nucleotide archive. *Nucleic Acids Res*. 2010;39(Suppl.1):D28-D31

doi 10.18699/bgrs2024-06-18

Гормональный статус и его регуляция на начальных этапах роста растений *Solanum tuberosum*, экспрессирующих *Bt*-ген

Пузина Т.*, Король В., Макеева И.
Орловский государственный университет им. И.С. Тургенева, Орёл, Россия
** tipuzina@gmail.com*

Ключевые слова: трансгенные растения с *Bt*-геном; фитогормоны; микроэлементы цинк и бор; *Solanum tuberosum*

Мотивация и цель: В фундаментальных исследованиях все чаще используют генно-инженерные технологии, которые способствуют выявлению молекулярно-генетических механизмов физиолого-биохимических процессов в растительном организме. Известно, что фитогормонам принадлежит ведущая роль в регуляции роста и развития растений. Вместе с тем изучению гормонального статуса трансгенных растений в отечественных и зарубежных исследованиях уделяется недостаточно внимания. Имеются данные о влиянии экзогенных гормонов на ростовые реакции трансгенных растений [1, 2]. Лишь в отдельных работах показано изменение содержания одной или двух групп фитогормонов у трансформантов [3, 4]. Однако проявление физиологического ответа зависит от той гормональной ситуации, а именно содержания и соотношения фитогормонов, которая складывается на определенном этапе онтогенеза. При этом с практической точки зрения важно наметить пути регуляции гормонального статуса трансгенных растений. В настоящее время широкое распространение получили растения, трансформированные *Bt*-геном из *Bacillus thuringiensis*, экспрессирующие белки-эндотоксины, губительно действующие на жесткокрылых, и в частности на колорадского жука [5]. При этом основное внимание авторы уделяют экологическим аспектам возделывания данных культур. Не найдено работ, в которых исследуется реакция гормональной системы растения на трансформацию *Bt*-геном, что не позволяет наметить пути ее регуляции. Целью работы было изучение содержания абсцизовой кислоты, зеатина, индолилуксусной кислоты у исходных и трансформированных *Bt*-геном растений картофеля на начальном этапе роста почек апикальных глазков клубней. Одновременно исследовали влияние незаменимых микроэлементов цинка и бора на содержание фитогормонов.

Методы и алгоритмы: Объектом исследования служили растения картофеля (*Solanum tuberosum* L.) сорта Супериор и полученные на их основе компанией Monsanto трансгенные растения с *Bt*-геном из почвенной бактерии *Bacillus thuringiensis*, продуцирующей дельта-эндотоксины. Клубни трансгенов были предоставлены лабораторией генетики ВНЦ КХ им. А.Г. Лорха. Опыты проводили с клубнями через 72 часа после перенесения их из овощехранилища (+4 °C) в условия лаборатории (20 ± 2 °C). Клубни проращивали в кюветах с увлажненными опилками. Варианты опыта включали замачивание клубней исходных и трансгенных растений на 6 часов в растворах $ZnSO_4$ ($3 \cdot 10^{-3}$ М), H_3BO_3 ($8 \cdot 10^{-3}$ М), контрольных – в воде. Анализировали апикальные глазки клубней (почка-глазок с коровой паренхимой) через 72 часа от начала прорастания. Содержание

фитогормонов ауксинов, цитокининов и абсцизовой кислоты определяли методом твердофазного иммуноферментного анализа [6] с использованием реактивов фирмы «Уралинвест» (Уфа). В качестве стандартных растворов фитогормонов были взяты ИУК, зеатин, АБК (Serva, Германия). На рис. 1 и в табл. 1 представлены средние арифметические из пяти биологических и трех аналитических повторностей. Достоверность оценивали с помощью критерия Стьюдента, считая достоверными различия при уровне доверительной вероятности выше 0.95.

Результаты: Через трое суток проращивания клубней глазки трансгенов содержали в 1.9 раза больше абсцизовой кислоты (рис. 1, а). Обогащение клубней цинком уменьшило содержание данного гормона у глазков исходных и трансгенных растений в равной степени – в 1.7 раза. Однако микроэлемент бор не изменил количество АБК у трансгенов, но снизил у исходных растений в 1.6 раза. В отличие от АБК, трансгены содержали несколько меньше цитокининов (на 17 %) по сравнению с глазками клубней исходных растений (см. рис. 1, б). Обогащение цинком значительно повысило уровень зеатина – в 2.3 раза у исходных и в 3 раза у трансгенов. Обработка микроэлементом бором была менее эффективна по сравнению с цинком. Ауксины были более чувствительны, чем цитокинины, к трансформации растений *Bt*-геном: их содержание было ниже исходных в 2.5 раза (см. рис. 1, в). Микроэлементы повысили уровень ИУК, причем в большей степени у трансгенов (цинк – в 2.2, бор – в 2.5 раза), тогда как у нетрансформантов цинк – на 25 %, бор – на 76 %.

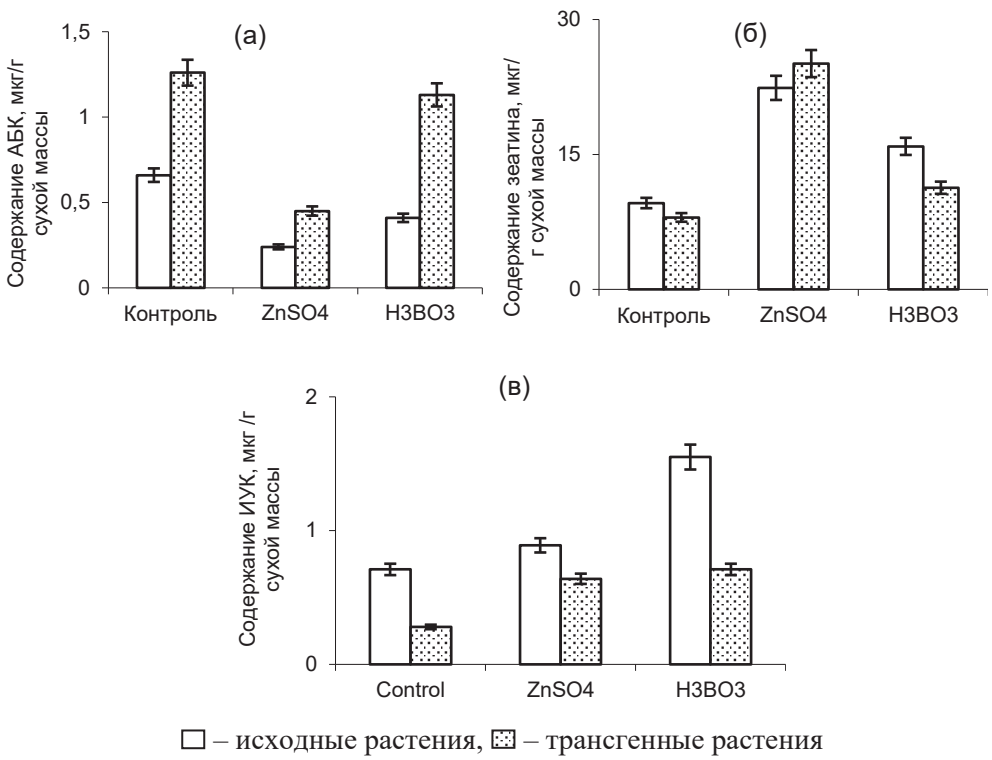


Рис. 1. Влияние микроэлементов на содержание фитогормонов в глазках клубней исходных и трасгенных растений: а – АБК; б – зеатин, в – ИУК

Анализ соотношения содержания фитогормонов стимуляторов (акусины+ цитокинины) к АБК свидетельствует о значительно меньшем соотношении у трансформантов за счет преобладания абсцизовой кислоты (см. табл. 1). Обогащение микроэлементами увеличило данное соотношение вне зависимости от трансформации. Более эффективным был микроэлемент цинк.

Таблица 1. Влияние микроэлементов на соотношение фитогормонов в глазках клубней исходных и трансгенных растений

Вариант	Зеатин + ИУК/АБК	
	Исходные растения	Трансгенные растения
Контроль	17.2	6.6
$ZnSO_4$	9.70	57.2
H_3BO_3	41.8	9.7

Выводы: Таким образом, проведенное исследование свидетельствует, что гормональная система *Solanum tuberosum* чувствительна к трансформации *Bt*-геном. Глазки клубней трансформантов в первые трое суток прорастания характеризуются повышенным содержанием АБК, но меньшим – зеатина и ИУК. Обработка клубней сернокислым цинком и борной кислотой снижает количество АБК, но повышает цитокинины и ауксины вне зависимости от трансформации. Наибольший положительный эффект на соотношение гормонов-стимуляторов к АБК оказал микроэлемент цинк как за счет значительного уменьшения АБК, так и за счет повышения уровня зеатина.

Hormonal status and its regulation at the initial stages of growth of *Solanum tuberosum* plants expressing the *Bt* gene

Puzina T.*, Korol V., Makeeva I.

Orel State University named after I.S. Turgenev, Orel, Russia

** tipuzina@gmail.com*

Key words: transgenic plants with the *Bt* gene; phytohormones; microelements zinc and boron; *Solanum tuberosum*

Motivation and Aim: In fundamental research, genetic engineering technologies are increasingly being used, which help to identify the molecular genetic mechanisms of physiological and biochemical processes in the plant organism. It is known that phytohormones play a leading role in the regulation of plant growth and development. At the same time, insufficient attention is paid to the study of the hormonal status of transgenic plants in domestic and foreign studies. There is data on the influence of exogenous hormones on the growth responses of transgenic plants [1, 2]. Only a few studies have shown changes in the content of one or two groups of phytohormones in transformants [3, 4]. However, the manifestation of a physiological response depends on the hormonal situation, namely, the content and ratio of phytohormones, which develops at a certain stage of ontogenesis. At the same time, from a practical point of view, it is important to outline ways to regulate the hormonal status of transgenic plants. Currently, plants transformed with the *Bt* gene from *Bacillus thuringiensis*, expressing endotoxin proteins that have a detrimental effect on beetles and, in particular, the Colorado potato

beetle, are widely used [5]. At the same time, the authors pay main attention to the environmental aspects of the cultivation of these crops. No works have been found that study the reaction of the plant hormonal system to transformation by the *Bt* genome, which does not allow us to outline ways of its regulation. The purpose of the work was to study the content of abscisic acid, zeatin, and indolylacetic acid in the original and transformed potato plants with the Bt gene at the initial stage of growth of the buds of the apical eyes of tubers. At the same time, the influence of the essential microelements zinc and boron on the content of phytohormones was studied.

Methods and Algorithms: The object of the study was potato plants (*Solanum tuberosum* L.) of the Superior variety and transgenic plants obtained on their basis by Monsanto with the Bt gene from the soil bacterium Bacillus thuringiensis, which produces delta-endotoxins. Transgene tubers were provided by the Laboratory of Genetics of the All-Russian Scientific Center for Chemistry named after A.G. Lorch. Experiments were carried out with tubers 72 hours after they were transferred from the vegetable storehouse (+4 °C) to laboratory conditions (20 ± 2 °C). Tubers were germinated in ditches with moistened sawdust. Variants of the experiment included soaking the tubers of the original and transgenic plants for 6 hours in solutions of $ZnSO_4$ ($3 \cdot 10^{-3}$ M), H_3BO_3 ($8 \cdot 10^{-3}$ M), control ones – in water. The apical eyes of tubers (bud-eye with cortical parenchyma) were analyzed 72 hours from the beginning of germination. The content of phytohormones auxins, cytokinins and abscisic acid was determined by enzyme-linked immunosorbent assay [6] using reagents from Uralinvest (Ufa). IAA, zeatin, and ABA (Serva, Germany) were taken as standard solutions of phytohormones. The figure and table show the arithmetic means of five biological and three analytical replicates. Reliability was assessed using the Student's test, considering differences at a confidence level above 0.95 to be significant.

Results: After three days of tuber germination, the transgene eyes contained 1.9 times more abscisic acid (Fig. 1a). Enrichment of tubers with zinc reduced the content of this hormone in the eyes of the original and transgenic plants to the same extent – 1.7 times. However, the microelement boron did not change the amount of ABA in transgenes, but decreased it by 1.6 times in the original plants. In contrast to ABA, the transgenes contained slightly less cytokinins (by 17 %) compared to the tuber eyes of the original plants (Fig. 1b). Zinc enrichment significantly increased the level of zeatin – 2.3 times in the original and 3 times in the transgenes. Treatment with the trace element boron was less effective compared to zinc. Auxins were more sensitive than cytokinins to the transformation of plants with the Bt gene, namely, their content was 2.5 times lower than the initial ones (Fig. 1c). Microelements increased the level of IAA, and to a greater extent in transgenes (zinc – by 2.2, boron – by 2.5 times), while in non-transformants zinc – by 25 %, boron – by 76 %.

Analysis of the ratio of the content of stimulant phytohormones (acousins + cytokinins) to ABA indicates a significantly lower ratio in transformants due to the predominance of abscisic acid (Table 1). Micronutrient enrichment increased this ratio regardless of transformation. The microelement zinc was more effective.

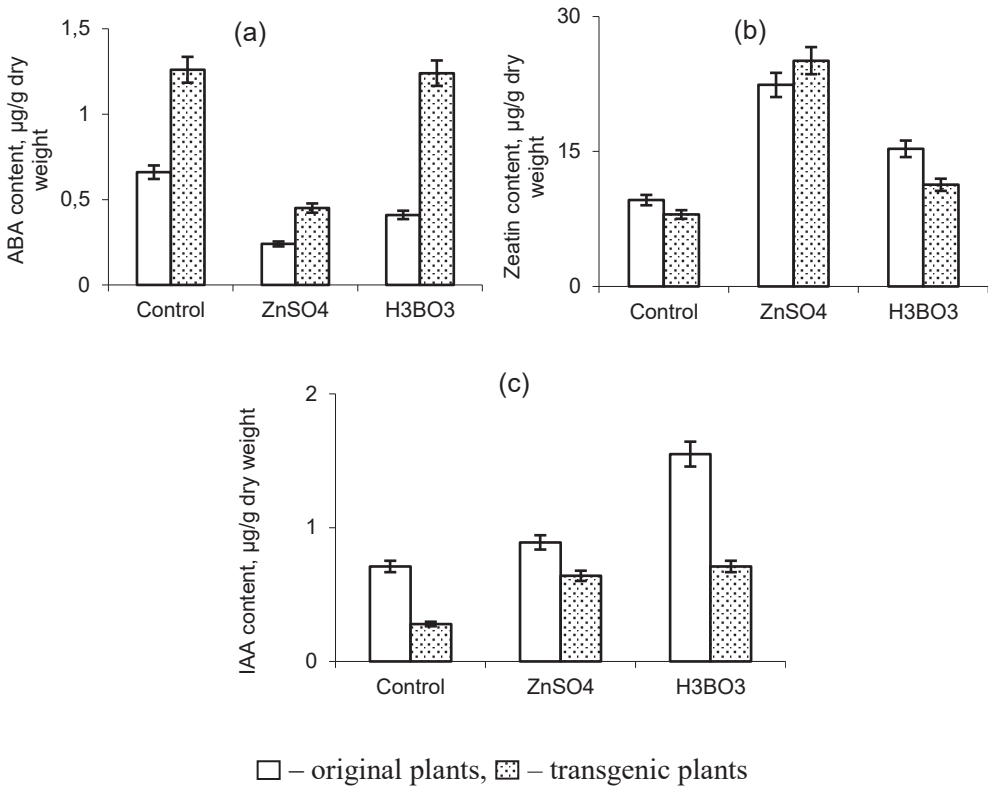


Fig. 1. The influence of microelements on the content of phytohormones in the eyes of tubers of original and transgenic plants: a – ABA; b – zeatin, c – IAA

Table 1. The influence of microelements on the ratio of phytohormones in the eyes of tubers of original and transgenic plants

Option	Zeatin + IAA/ABA	
	Original plants	Transgenic plants
Control	17.2	6.6
$ZnSO_4$	9.70	57.2
H_3BO_3	41.8	9.7

Conclusion: Thus, the study indicates that the hormonal system of *Solanum tuberosum* is sensitive to transformation by the *Bt* gene. The eyes of transformant tubers in the first three days of germination are characterized by an increased content of ABA, but less zeatin and IAA. Treatment of tubers with zinc sulfate and boric acid reduces the amount of ABA, but increases cytokinins and auxins, regardless of transformation. The microelement zinc had the greatest positive effect on the ratio of stimulant hormones to ABA due to both a significant decrease in ABA and an increase in zeatin levels.

Список литературы/References

1. Додуева И.Е., Фролова Н.В., Власенко М.А., Монахова В.А., Лутова Л.А. Трансформация инбредных линий редиса (*Raphanus sativus* L.) генами т-ДНК агробактерий: изменение опухолевого фенотипа и реакции на фитогормоны у трансгенных растений. *Вестник биотехнологии и физико-химической биологии им. Ю.А. Овчинникова*. 2005;1(2):22-29
   [Dodueva I.E., Frolova N.V., Vlasenko M.A., Monakhova V.A., Lutova L.A. Transformation of inbred lines of radish (*Raphanus sativus* L.) with t-DNA genes of agrobacteria: changes in tumor phenotype and reaction on phytohormones in transgenic plants. *Vestnik Biotekhnologii I Fiziko-Khimicheskoy Biologii im. Yu.A. Ovchinnikova*. 2005;1(2):22-29 (in Russian)]
2. Кулуев Б.Р., Сафиуллина М.Г., Князев А.В., Никоноров Ю.М., Чемерис А.В. Роль гена NtEXP A6 в регуляции роста органов табака. *Биомика*. 2014;6(1):1-12
   [Kuluev B.R., Safiullina M.G., Knyazev A.V., Nikonorov Yu.M., Chemeris A.B. The role of the NtEXP A6 gene in the regulation of tobacco organ growth. *Biomika*. 2014;6(1):1-12 (in Russian)]
3. Пак Чун Ир. Изучение влияния повышенных уровней фитогормонов в трансгенных растениях на экспрессию хлоропластных генов. Автореферат дис. … кандидата биологических наук. Москва, 1991
   [Pak Chun Ir. Study of the influence of increased levels of phytohormones in transgenic plants on the expression of chloroplast genes. Abstract of thesis … candidate of biological sciences. Moscow, 1991 (in Russian)]
4. Аксенова Н.П., Константинова Т.Н., Ложникова В.Н., Голяновская С.А., Гукасян И.А., Гатс К., Романов Г.А. Фотопериодическая и гормональная зависимость клубнеобразования у картофеля, трансформированного геном PHYB Arabidopsis. *Физиология растений*. 2005;52(5):701-707
   [Aksenova N.P., Konstantinova T.N., Lozhnikova V.N., Golyanovskaya S.A., Gukasyan I.A., Gats K., Romanov G.A. Photoperiodic and hormonal dependence of tuber formation in potatoes transformed with the Arabidopsis PHYB gene. *Fiziologiya Rasteniy*. 2005;52(5):701-707 (in Russian)]
5. Викторов А.Г. Эколого-физиологические особенности Bt-растений, приводящие к вспышкам численности вторичных вредителей. *Физиология растений*. 2017;64(4):243-250
   [Viktorov A.G. Ecological and physiological characteristics of Bt plants leading to outbreaks of secondary pests. *Fiziologiya Rasteniy*. 2017;64(4):243-250 (in Russian)]
6. Веселов С.Ю. Использование антител для количественного определения, очистки и локализации регуляторов роста растений. Уфа: БГУ, 1998
   [Veselov S.Yu. Use of antibodies for quantification, purification and localization of plant growth regulators. Ufa: BSU, 1998 (in Russian)]

doi 10.18699/bgrs2024-06-19

Внутри- и межвидовые вариации длины теломер Мохообразных

Санникова А.В.[1]*, Шарипова М.Р.[1], Шакиров Е.В.[1, 2], Валеева Л.Р.[1, 2]

[1] *Казанский (Приволжский) федеральный университет, Казань, Россия*

[2] *Department of Biological Sciences, College of Science, Marshall University, USA*

** anastasya.sannikova@bk.ru; AnVSannikova@stud.kpfu.ru*

Ключевые слова: теломеры; мхи; печеночники; *Ceratodon purpureus*; *Marchantia polymorpha*; *Sphagnum*; *Physcomitrium patens*

Мотивация и цель: Теломеры представляют собой нуклеопротеиновые структуры на физических концах линейных хромосом эукариот, участвующие в защите ДНК от повреждений и определяющие жизнеспособность клетки. Строение теломер эволюционно консервативно в разных царствах живых организмов, включая животных, растения и грибы. Так и для большинства видов растений характерна теломерная ДНК, состоящая из повторов $(TTTAGGG)_n$. Однако средняя длина теломер может быть не только видоспецифична; существенные различия могут наблюдаться даже в пределах популяций одного вида.

Основной модельной системой для изучения биологии теломер растений на протяжении последних десятилетий было небольшое цветковое растение *Arabidopsis thaliana*. Хотя это модельное растение вносит колоссальный вклад в исследование гомеостаза теломер, многие вопросы о биологии теломер растений до сих пор остаются открытыми. В связи с этим использование новых растительных модельных организмов в исследованиях биологии теломер представляет большой интерес. Мохообразные являются одной из перспективных групп для изучения молекулярной генетики и эволюции растений, в частности биологии теломер, поскольку их можно рассматривать и как альтернативную, и как аналогичную ветвь эволюции наземных растений. Кроме того, мохообразные обладают рядом преимуществ, таких как: низкая избыточность генома из-за отсутствия недавних дупликаций, преобладание гаплоидной фазы жизненного цикла, двудомность многих видов бриофитов. Таким образом, мохообразные в силу своего филогенетического положения являются новой модельной системой для изучения биологии теломер растений [1].

Целью данной работы был сравнительный анализ длины теломер у разных видов бриофитов, а также среди популяций одного вида.

Методы и алгоритмы: В работе использовали аксенические культуры мхов: *Physcomitrium patens*, 4 экотипа (Gransden, Великобритания; Reute, Германия; Villersexel, Франция; Kaskaskia, США); культуры *Ceratodon purpureus*: R40 (мужское растение, Нью-Йорк, США), GG1 (женское растение, Австрия), B150 и B190 (женское и мужское растение, Аляска, США), печеночник *Marchantia polymorpha* Takaragaike-1 (мужское растение) и Takaragaike- 2 (женское растение; Япония), а также природные изоляты сфагнума: *Sphagnum fallax* MW (Миннесота, США), *Sphagnum girgensohnii* (Свердловская область, Россия) и *Sphagnum* sp. (Республика Марий Эл, Россия). Анализ длины теломер проводили с помощью TRF-анализа (анализ терминальных рестрикционных фрагментов) совместно с

Саузерн-блотом [2]. Среднюю длину теломер рассчитывали с помощью программы TeloTool [3]. Статистический анализ проводили с помощью программы GraphPad Prism v.8.

Результаты: Мы показали, что разные экотипы растений *P. patens* имеют длину теломер в диапазоне от 1.08 ± 0.13 т.п.н. (Kaskaskia) до 1.71 ± 0.19 т.п.н. (Villersexel), длина теломер у экотипа Reute составила 1.29 ± 0.04 т.п.н., у экотипа Gransden – 1.21 ± 0.06 т.п.н. Таким образом, длина теломер *P. patens* оказалась в 1.5–3 раза короче длины теломер модельного покрытосеменного растения *A. thaliana* экотип Col-0 (в среднем 2.5–4.5 т.п.н.) [4]. Согласно полученным результатам, длина теломер мха *P. patens* варьирует между экотипами, при этом теломеры экотипа Kaskaskia являются самыми короткими, а теломеры экотипа Villersexel – самыми длинными. Кроме того, мы показали, что все изученные экотипы имеют специфические теломерные последовательности, предположительно внутрихромосомной локализации, различающиеся внутри экотипов по расположению и длине.

Далее мы проанализировали длину теломер у четырех природных изолятов мха *C. purpureus*. Средние значения длины теломер *C. purpureus* варьировали от самых коротких 0.68 ± 0.04 т.п.н. в линии GG1 до самых длинных 1.15 ± 0.14 т.п.н. в линии B190. Изолят B150 имел более длинные теломеры ($p = 0.03$), чем изолят B190, а также длиннее, чем линия R40 ($p = 0.002$). Однако выводы о корреляции длины теломер с полом растений требуют дальнейшего изучения.

Мы сравнили распределение длин теломер у печеночника *M. polymorpha*. TRF-анализ показал, что среднее значение длины теломер у мужского растения (Так-1) составляет 2.15 ± 0.25 т.п.н., у женского растения (Так-2) длина теломер составляет 2.45 ± 0.25 т.п.н. Женская линия имела более длинные теломеры. Тем самым было показано, что печеночник *M. polymorpha* обладает наиболее длинными теломерами среди изученных бриофитов *P. patens* и *C. purpureus*.

Далее мы определили длины теломер у мхов рода *Sphagnum*. TRF-анализ показал вариабельность в длине теломер у трех изолятов мха *Sphagnum*. Средняя длина теломер у *S. fallax* составила 1.86 ± 0.08 т.п.н.; у вида *S. girgensohnii* – 1.56 ± 0.21 т.п.н., у изолята *Sphagnum* sp. – 1.35 ± 0.11 т.п.н., что также указывает на естественную вариабельность длины теломер между разными видами. Также у *S. fallax* был обнаружен уникальный внутрихромосомный теломерный повтор (ITS) размером ~ 2.1 т.п.н., тогда как в других изолятах ITS размером около 2 т.п.н. практически не выявлялся. Таким образом, для мхов рода *Sphagnum* характерно разнообразие длины теломер и внутрихромосомных теломерных повторов.

Нами была также изучена динамика изменений длины теломер в клетках протонемы *P. patens* (экотип Reute) и *C. purpureus* GG1 и R40, выращенных в течение 14, 28 и 42 дней. В результате было показано, что длина теломер остается неизменной в активно делящихся вегетативных тканях протонемы на протяжении периода культивирования, что позволяет предположить, что множественные деления клеток, связанные с быстрым ростом протонемы, не влияют на длину теломер. Кроме того, такая же длина теломер наблюдалась в гаметофорах растений *P. patens*.

Выводы: Таким образом, мы показали, что все проанализированные мохообразные обладают относительно короткими теломерами, в среднем ниже 2.5 т.п.н., а в некоторых случаях ниже 1.5 т.п.н., что, в свою очередь, нехарактерно для большинства цветковых растений. Длина теломер мохообразных может

варьировать как между видами, так и внутри одного вида. Кроме того, в геномной ДНК всех изученных мохообразных были обнаружены последовательности ITS. Значительная вариабельность длины теломер была показана для четырех проанализированных изолятов мха *C. purpureus*. Также выявлены различия в длине теломер между мужским и женским растениями *M. polymorpha*. Однако для подтверждения корреляции длины теломер с полом у мохообразных необходимо проведение дальнейших исследований. Помимо того, было показано, что длина теломер остается стабильной в течение культивирования на разных стадиях вегетативного роста изученных линий растений. Таким образом, наблюдаемые внутривидовые различия в длине теломер у бриофитов могут стать основанием для их будущего использования в картировании и количественных исследованиях локусов генов, ассоциированных с длиной теломер, и дальнейшего обнаружения новых регуляторных генов биологии теломер растений.
Финансирование: Работа выполнена на технической базе программы стратегического академического лидерства «Приоритет-2030».

Intra- and interspecific variations in the telomere length of bryophytes

Sannikova A.V.[1]*, Sharipova M.R.[1], Shakirov E.V.[1, 2], Valeeva L.R.[1, 2]

[1] *Kazan Federal University, Kazan, Russia*

[2] *Department of Biological Sciences, College of Science, Marshall University, USA*

* *AnVSannikova@stud.kpfu.ru; anastasya.sannikova@bk.ru*

Key words: telomeres; mosses; liverworts; *Ceratodon purpureus*; *Marchantia polymorpha*; *Sphagnum*; *Physcomitrium patens*

Motivation and Aim: Telomeres are nucleoprotein structures at the physical ends of eukaryotic chromosomes playing a crucial role in protecting DNA from damage and in cell fate. Telomere architecture is evolutionarily conserved across different organisms including animals, plants, and fungi Most plant species have telomeres that consist of TTTAGGG repeats. However, the average length of telomers is not only species-specific, but its substantial variation could be found within natural populations of species.

Historically, the main model system to study plant telomere biology was the small flowering plant *Arabidopsis thaliana*. Although this model plant provides a great contribution to numerous studies in the telomere homeostasis, many questions about the biology of plant telomeres remain unanswered. Therefore, the usage of new plant model organisms in the telomere biology studies are of the great interest due to the limited data in this field. Bryophytes are one of the promising plant groups for the studies of molecular genetics and evolution of plants, particularly telomere biology, because they could be considered both as the alternative and as the analogous evolution branch of the land plants. In addition, bryophytes have several advantages, such as: low genetic redundancy due to the absence of recent genome duplications, the predominant phase of lifecycle for all bryophytes is haploid, many bryophyte species are dioecious. Thus, due to their phylogenetic position, bryophytes are a new model system for studying plant telomere biology.

The aim of the current work was the comparative analysis of telomere length both among different bryophyte species and within the same species.

Methods and Algorithms: In the study we used axenic cultures of mosses: *Physcomitrium patens*, 4 ecotypes (Gransden, UK; Reute, Germany; Villersexel, France; Kaskaskia, USA); *Ceratodon purpureus* cultures: R40 (male plant line, New York, USA), GG1 (female plant line, Austria), B150 and B190 (female and male lines, Alaska, USA), a liverwort *Marchantia polymorpha* (cultivars Takaragaike-1 (male) and Takaragaike-2 (female), Japan), and natural isolates of peatmoss *Sphagnum fallax* MW (MN, USA), *Sphagnum girgensohnii* (Sverdlovsk Oblast, Russia) and *Sphagnum* sp. (Republic of Mari-El, Russia). Telomere length analysis was performed by TRF analysis (Terminal Restriction Fragment analysis) coupled with Southern blot [1]. The average telomere length was calculated using the TeloTool software [2]. Statistical analysis was carried out by GraphPad Prism v.8 software.

Results: We have shown that different ecotypes of *P. patens* plants have a telomere length in the range from 1.08 ± 0.13 kb (Kaskaskia) to 1.71 ± 0.19 kb (Villersexel), Telomere length in the ecotypes of Reute being 1.29 ± 0.04 kb. Same as in the Gransden being 1.21 ± 0.06 kb. It was 1.5–3 times shorter than the telomere length of the model angiosperm plant *A. thaliana* Col-0 (2.5–4.5 kb in the average) [3]. Our data indicate that telomere length varies between analyzed *P. patens* ecotypes, with Kaskaskia telomeres being the shortest and Villersexel telomeres being the longest. In addition, we have shown that all the ecotypes studied have specific telomeric sequences (ITS), presumably of intra-chromosomal localization, and differing in their location and length within the ecotypes. We analyzed telomere length in four natural isolates of *C. purpureus*. Mean TRF values in *C. purpureus* accessions varied from the shortest 0.68 ± 0.04 kb in the GG1 ecotype to the longest 1.15 ± 0.14 kb in the B190 accession. B150 had longer telomeres ($p = 0.03$) than the male Alaskan isolate B190, as well as longer than the other male line R40 ($p = 0.002$). We did not find sex-specific correlations in mean TRF values between female and male lines.

We compared telomere lengths distribution in strains of liverwort *M. polymorpha*. TRF analysis of *M. polymorpha* telomeres indicated that mean TRF values in male (Так-1) 2.15 ± 0,25 kb and female (Так-2) 2.45 ± 0.25 kb lines were different. The female strain having longer telomeres. We note that telomeres in the liverwort *M. polymorpha* are longer than in all analyzed ecotypes of the model mosses *P. patens* and *C. purpureus*.

We next analyzed telomere lengths of *Sphagnum* species. TRF analysis revealed variable telomere length in the three *Sphagnum* isolates. Mean telomere lengths differed in *S. fallax* 1.86 kb ± 0.08 kb; *S. girgensohnii* 1.56 ± 0.21 kb and *Sphagnum* sp. 1.35 ± 0.11 kb, implying natural variation in telomere length between different. A strong signal of intrachromosomal telomeric repeats was identified in *S. fallax* MW DNA (~ 2.1 kb), which was not nearly as strong in the other two isolates. Therefore, it was shown that there are intraspecies variations in telomere length of *Sphagnum* and also differences in the telomeric repeats location on chromosomes.

We evaluated telomere length dynamics of *P. patens* protonema (Reute ecotype) and *C. purpureus* GG1 and R40 protonema tissues grown on plates for 14, 28, and 42 days. We observed no mean TRF changes in vegetatively grown protonema tissue for the entirety of the cultivation suggesting no telomere length change due to multiple cell divisions associated with the rapid growth of protonema. Furthermore, the same telomere length was observed in 2-month-old gametophores (*P. patens*).

Conclusion: Therefore, we have shown that the mean telomere length in all analyzed bryophytes is relatively short, commonly below 2.5 kb and below 1.5 kb in some cases, which is unusual results comparing to the flowering plants. It was also shown that the telomere length of bryophytes can vary both between species and within the same species. Additionally, ITS sequences were detected in genomic DNA of all studied bryophytes. In four analyzed *C. purpureus* isolates, all lines showed substantial variation in the telomere length. Similarly, telomere length differences in the two tested male and female accessions of *M. polymorpha* were also identified. However, the correlation with sex in these bryophytes remains to be studied. We note that telomere length remains relatively stable over time in different tissues of vegetatively growing plants. Thus, the observed substantial intra-species variation in telomere length can serve as a strong foundation for their future use in association mapping and quantitative trait loci studies to discover causal genetic variants.

Funding: The study was carried out thanks to the technical resources of the Strategic Academic Leadership Program of the Kazan Federal University (PRIORITET-2030).

Список литературы/References

1. Valeeva L.R. et al. Telomere length variation in model bryophytes. *Plants (Basel)*. 2024;13(3):387. doi 10.3390/plants13030387
2. Nigmatullina L.R. et al. Non-radioactive TRF assay modifications to improve telomeric DNA detection efficiency in plants. *Bionanoscience*. 2016;6(4):325-328
3. Abdulkina L.R. et al. Comparative application of terminal restriction fragment analysis tools to large-scale genomic assays. *Int J Mol Sci*. 2023;24(24):17194
4. Shakirov E.V., Shippen D.E. Length regulation and dynamics of individual telomere tracts in wild-type Arabidopsis. *Plant Cell*. 2004;16(8):1959-1967

doi 10.18699/bgrs2024-06-20

Регуляция активности гена *MAKR6* в корне *Arabidopsis thaliana* L.

Сидоренко А.Д.[1, 2]*, Новикова Д.Д.[3], Миронова В.В.[1, 2], Землянская Е.В.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Университет Лозанны, Лозанна, Швейцария*

* *a.sidorenko1@g.nsu.ru*

Ключевые слова: корень; сосудистая система; ауксин; цитокинин; мембран-ассоциированные регуляторы киназ

Мотивация и цель: Мембран-ассоциированные регуляторы киназ (MAKR) – семейство, включающее семь белков [1]. Его представители являются важными регуляторами развития растений, но неизученным остается ген *MAKR6*. Ранее на основании исследования паттерна транскрипционной активности *MAKR6* в тканях проростка *Arabidopsis thaliana* мы показали, что этот ген может быть новым тканеспецифическим ауксин-чувствительным регулятором развития растений [2]. Регуляция транскрипции играет ключевую роль в создании определенного паттерна работы гена [3], и у растений она часто контролируется фитогормонами. Целью нашей работы является исследование механизмов регуляции активности гена *MAKR6* в корне *Arabidopsis thaliana* L.

Материалы и алгоритмы: Измерение уровня экспрессии в ответ на обработку фитогормонами проводили с помощью количественной ПЦР с обратной транскрипцией (ОТ-ПЦР). Список транскрипционных факторов (ТФ), являющихся потенциальными регуляторами *MAKR6*, был составлен на основании публично доступных данных DAP-seq [4]. Репортерные линии *pMAKR6:nls3GFP* на фоне мутации *mp/arf5* были получены путем скрещивания репортерных и мутантных линий. Анализ паттерна GFP в тканях репортерных линий проводили с помощью эпифлуоресцентной и конфокальной микроскопии.

Результаты: С помощью количественной ОТ-ПЦР мы показали, что *MAKR6* повышает свою экспрессию в корне уже после тридцатиминутной обработки ауксином. Активация гена в ответ на обработку цитокинином наблюдается через 24 часа и предположительно является результатом вторичного ответа на фитогормон. В промоторе *MAKR6* были найдены потенциальные сайты связывания ТФ семейства ARF, отвечающих за первичный ответ на ауксин, а также ряда цитокинин-чувствительных ТФ, локализованные в пиках DAP-seq для соответствующих ТФ. Транскрипционный ответ гена *MAKR6*, а также репортерной конструкции *pMAKR6:nls3GFP* на ауксин на фоне мутации по гену *MP/ARF5* (ключевой регулятор развития сосудистой системы корня) не отличался от такового в генетическом окружении дикого типа. Вероятно, чувствительность *MAKR6* к ауксину регулируется другими представителями семейства ARF. Также показано, что цитокинин-чувствительный регуляторный элемент располагается за пределами проксимальной области промотора.

Выводы: На основании полученных результатов можно сказать, что транскрипция *MAKR6* в корне регулируется ауксином и цитокинином, которые известны как

ключевые регуляторы формирования и поддержания бисимметричной организации сосудистой системы в кончике корня.
Финансирование: Работа поддержана грантом РНФ № 20-14-00140.

Regulation of *MAKR6* gene activity in *Arabidopsis thaliana* L. root

Sidorenko A.D.[1, 2]*, Novikova D.D.[3], Mironova V.[1, 2], Zemlyanskaya E.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *University of Lausanne, Lausanne, Switzerland*

* *a.sidorenko1@g.nsu.ru*

Key words: root; vascular tissue; auxin; cytokinin; membrane-associated kinase regulator

Motivation and Aim: Membrane-associated kinase regulators (MAKR) family contains seven proteins, most of which are important plant development regulators [1]. *MAKR6* is the only family member, whose functions are poorly understood. Previously, based on the patterns of *MAKR6* transcriptional activity in Arabidopsis seedling tissues, we demonstrated that *MAKR6* might be a new tissue-specific auxin-sensitive regulator of plant development [2]. Regulation of transcription plays a pivotal role in gene expression patterning [3], and, in plants, it is often controlled by phytohormones. Here, we investigate the mechanisms of *MAKR6* transcriptional regulation in *Arabidopsis thaliana* L. root.
Methods and Algorithms: We measured the overall expression level of *MAKR6* using RT-qPCR. The list of transcription factors (TF), which are the potential regulators of *MAKR6* expression, was compiled based on publicly available DAP-seq data [4]. The reporter lines, which express *pMAKR6:nls3GFP* in the *mp/arf5* background were generated by crossing the corresponding *A. thaliana* reporter lines and mutant lines. To analyze GFP pattern in the tissues of the reporter lines, we used epifluorescent and confocal microscopy techniques.
Results: Using RT-qPCR, we demonstrated that *MAKR6* is upregulated after 30 minutes of auxin application. Gene activation by cytokinin application is observed after 24 hours, and, supposedly, is the result of a secondary response to the phytohormone. *MAKR6* promoter contains potential binding sites the ARF family TFs, which promote early response to auxin, as well as a number of cytokinin-sensitive TFs, localized in DAP-seq peaks for the corresponding TFs. The transcriptional response of both *MAKR6* gene and *pMAKR6:nls3GFP* reporter construct to auxin in the mutant background *mp/arf5* (ARF5 is known as a key regulator of the development of the root vascular system) did not differ from that in the wild-type background. Thus, *MAKR6* sensitivity to auxin is probably regulated by other ARF family members. We also demonstrated that the cytokinin-sensitive regulatory elements are located outside of the proximal promoter region of *MAKR6*.
Conclusion: Our results show that *MAKR6* transcription in the root is regulated by auxin and cytokinin, which are known as key regulators of the formation and maintenance of the bisymmetric organization of the vascular system in the root tip.
Funding: This work was supported by the RSF grant No. 20-14-00140.

Список литературы/References

1. Novikova D.D., Korosteleva A.L., Mironova V., Jaillais Y. Meet your MAKR: the membrane-associated kinase regulator protein family in the regulation of plant development. *FEBS J.* 2022;289(20):6172-6186
2. Sidorenko A.D., Novikova D.D., Mironova V., Zemlyanskaya E. Understanding the role of MAKR6 in *Arabidopsis thaliana* L. root development. In: Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022). Novosibirsk, 2022;662
3. Lee J.Y., Colinas J., Wang J.Y., Mace D., Ohler U., Benfey P.N. Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. *PNAS*. 2006;103(15):6055-6060
4. O'Malley R.C., Huang S S.C., Song L., Lewsey M.G., Bartlett A., Nery J.R., Ecker J.R. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*. 2016;165(5):1280-1292

doi 10.18699/bgrs2024-06-21

Влияние организации целлюлозы на кислотоиндуцированное растяжение клеточных стенок в гипокотилях *Arabidopsis thaliana*

Суслов Д.В.
Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
d.suslov@spbu.ru

Ключевые слова: рост клеток растяжением; клеточная стенка; растяжимость клеточной стенки; целлюлоза; кислый рост; экспансины

Мотивация и цель: Высокая скорость увеличения объема растительных клеток в ходе роста растяжением достигается за счет накопления в них воды. Важнейшим биофизическим фактором, лимитирующим скорость роста клеток растений, является растяжимость клеточных стенок, которая зависит от трехмерной организации полимеров стенки и от активностей многочисленных ферментативных и неферментативных белков, непрерывно формирующих и разрушающих ковалентные и нековалентные связи между компонентами стенок. Центральную роль в регуляции растяжимости клеточных стенок играет целлюлоза – самый прочный компонент в данной структуре. Кислый рост является одним из ранних процессов в ходе индуцированного ауксином растяжения органов растений. Он связан с подкислением апопласта H+-АТФазами плазмалеммы, что активирует в клеточной стенке белки экспансины, увеличивающие ее растяжимость за счет разрушения водородных связей между структурными полимерами стенки [1]. Недавно продемонстрировали, что нарушение организации целлюлозы в клеточных стенках под влиянием ингибиторов и мутаций устраняет кислый рост гипокотилей арабидопсиса [2]. Поскольку эти данные не поддерживаются результатами некоторых наших экспериментов, мы изучили влияние разных типов реорганизации целлюлозы на кислый рост гипокотилей арабидопсиса. Исследовали эффекты изменения ориентации целлюлозы под влиянием фитогормона эпибрассинолида (ЭБЛ, 100 нМ) [3], ингибитора микротрубочек оризалина (250 нМ) и мутации pom2-4, уменьшения кристалличности целлюлозы под влиянием мутации ixr1-1 и снижения уровня целлюлозы в клеточных стенках в присутствии специфичного ингибитора изоксабена (0.1 нМ).
Методы и алгоритмы: Проявлением способности клеточных стенок поддерживать кислый рост является увеличение скорости их крипа (зависимая от времени необратимая деформация под влиянием постоянной нагрузки) в кислом буфере с pH 4.5–5.0 по сравнению с буфером с pH, близким к нейтральному (от 5.5 и выше), либо при сравнении растяжения в кислом буфере «нативных» клеточных стенок и клеточных стенок, инактивированных нагреванием, которое устраняет эндогенную активность экспансинов. В связи с этим крип клеточных стенок замороженных/оттаявших гипокотилей арабидопсиса измеряли экстенсометром, работающим при постоянной нагрузке, при pH 5, который активирует эндогенные экспансины клеточных стенок, и при pH 5 с тепловой инактивацией, т. е. при условиях, когда эндогенные экспансины не работают.

Результаты: Все экспериментальные обработки, использованные в данном исследовании, ингибировали рост побегов арабидопсиса: длина 4-суточных этиолированных гипокотилей уменьшалась в диапазоне от 46 % (оризалин) до 27 % (ЭБЛ). Варианты, в которых ориентация целлюлозы в клеточных стенках становилась более хаотичной (pom2-4, оризалин, ЭБЛ), сохраняли способность к кислому росту, но в разной степени. У мутанта pom2-4 скорость крипа клеточных стенок в кислом буфере увеличивалась по сравнению с контролем (Col-0) (рис. 1), тогда как оризалин (рис. 2) и ЭБЛ снижали скорость растяжения клеточных стенок *in vitro* у Col-0 при pH 5.

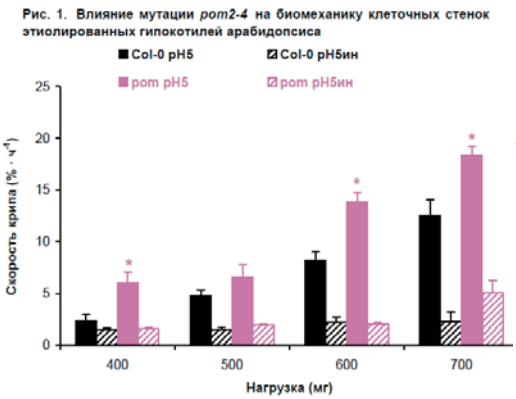


Рис. 1. Влияние мутации *pom2-4* на биомеханику клеточных стенок этиолированных гипокотилей арабидопсиса

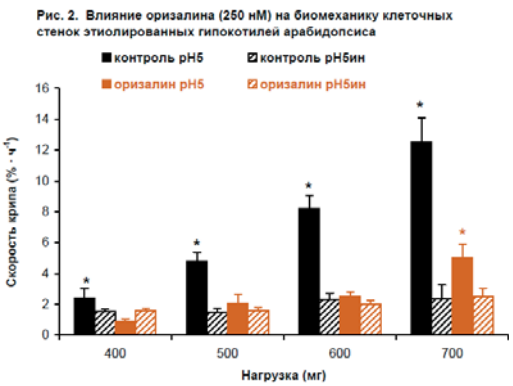


Рис. 2. Влияние оризалина (250 нМ) на биомеханику клеточных стенок этиолированных гипокотилей арабидопсиса

Такое различие эффектов может быть обусловлено неодинаковым влиянием указанных обработок на площадь поперечного сечения клеточных стенок гипокотилей, так как скорость крипа пропорциональна механическому напряжению (сила, поделенная на площадь поперечного сечения образца), генерированному в клеточных стенках постоянной нагрузкой. Обнаружили, что ЭБЛ достоверно увеличивает площадь поперечного сечения клеточных стенок. На основе этого рассчитали механическое напряжения в клеточных стенках в присутствии ЭБЛ и определили зависимость скорости крипа от напряжения. Характер данной зависимости демонстрирует, что эффект кислого буфера противоположен при низких и высоких значениях механического напряжения (рис. 3А).

Рис. 3А. Влияние ЭБЛ (100 нМ) на биомеханику нативных клеточных стенок этиолированных гипокотилей Col-0

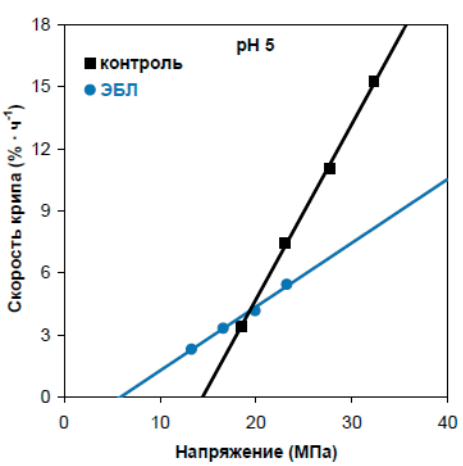
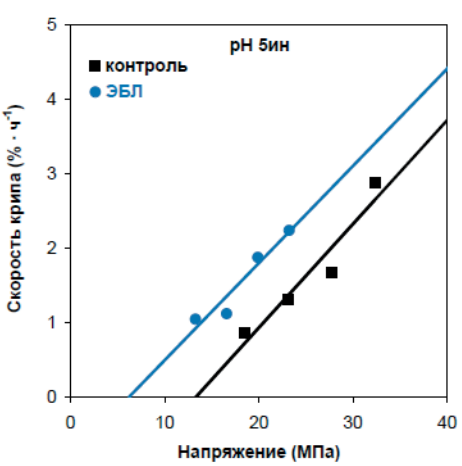


Рис. 3Б. Влияние ЭБЛ (100 нМ) на биомеханику инактивированных нагреванием клеточных стенок этиолированных гипокотилей Col-0



У мутанта ixr1-1, характеризующегося пониженной кристалличностью целлюлозы, был выявлен мощный индуцированный кислым буфером крип клеточных стенок, скорость которого была достоверно выше, чем у Col-0. Таким образом, уменьшение кристалличности микрофибрилл целлюлозы не является препятствием для кислого роста гипокотилей арабидопсиса.
В условиях ингибирования синтеза целлюлозы изоксабеном у растений Col-0 стимуляция крипа клеточных стенок кислым буфером сохранялась, но была существенно меньшей, чем у Col-0 в отсутствие ингибитора (см. рис. 1 и 2). Любопытно, что в присутствии изоксабена скорость крипа при pH 5 очень слабо изменялась при увеличении нагрузки (или механического напряжения), в отличие от контрольных растений дикого типа, у которых кислотоиндуцированное растяжение клеточных стенок значительно усиливалось при увеличении нагрузки (см. рис. 1). Этот факт косвенно свидетельствует об уменьшении активности эндогенных экспансинов под влиянием изоксабена. Не исключено, что в условиях ингибирования синтеза целлюлозы запускаются сигнальные каскады, активируемые нарушением целостности стенки [4], которые подавляют экспрессию экспансинов.
Выводы: Таким образом, в отличие от литературных данных [2], в настоящем исследовании было показано, что дезорганизация целлюлозы под влиянием фитогормонов, мутаций и ингибиторов не устраняет полностью, а по-разному модифицирует кислый рост гипокотилей арабидопсиса. Общий эффект может определяться как изменением структуры стенки, делающим ее более или менее оптимальной для работы экспансинов, так и изменением уровня экспрессии экспансинов с участием систем поддержания целостности клеточных стенок. Вклад последнего механизма можно было бы выяснить на основе данных транскриптомики и последующего биоинформационного анализа.
Финансирование: Исследование поддержано грантом РНФ (№ 23-24-00379).

Effects of cellulose arrangement on the acid-induced cell wall extension in *Arabidopsis thaliana* hypocotyls

Suslov D.V.*
St. Petersburg State University, St. Petersburg, Russia
** d.suslov@spbu.ru*

Key words: extension growth; cell wall; cell wall extensibility; cellulose; acid growth; expansins

Motivation and Aim: A high rate of increase in plant cell volume during extension growth is achieved due to water accumulation in them. The most important biophysical factor limiting plant cell growth rate is cell wall extensibility, which depends on the three-dimensional organization of wall polymers and on the activities of numerous enzymatic and non-enzymatic proteins that continuously form and destroy covalent and non-covalent bonds between the wall components. Cellulose, the strongest component in this structure, plays a central role in regulating the cell wall extensibility. The acid growth is one of the early processes during the auxin-induced elongation of plant organs. It is associated with the apoplast acidification by the plasmalemma H+-ATPases, which activates expansin proteins in the cell wall, increasing its extensibility due to the breakage of hydrogen bonds between the structural polymers of the wall [1]. It has

recently been demonstrated that cellulose disorganization in cell walls under the influence of inhibitors and mutations eliminates the acid growth of arabidopsis hypocotyls [2]. Since these data are not supported by the results of some of our experiments, we studied the effect of different types of cellulose reorganization on the acid growth of arabidopsis hypocotyls. The effects of a change in cellulose orientation under the influence of the phytohormone epibrassinolide (EBL, 100 nM) [3], the microtubule inhibitor oryzalin (250 nM) and the mutation *pom2-4*, a decrease in cellulose crystallinity under the influence of the mutation *ixr1-1* and a decrease in the level of cellulose in cell walls in the presence of a specific inhibitor isoxabene (0.1 nM) were studied.

Methods and Algorithms: A manifestation of the ability of cell walls to support the acid growth is an increase in their creep rate (the time-dependent irreversible deformation under the influence of a constant load) in an acidic buffer with a pH of 4.5–5.0 compared with a buffer with a pH close to neutral (from 5.5 and above), or when comparing the extension in an acidic buffer of "native" cell walls and heat-inactivated cell walls, which eliminates the activity of endogenous expansins. In this regard, the cell wall creep of frozen/thawed arabidopsis hypocotyls was measured with a constant-load extensometer, at pH 5, which activates endogenous cell wall expansins, and at pH 5 with heat-inactivation, i.e. under conditions when endogenous expansins do not work.

Results: All experimental treatments used in this study inhibited arabidopsis shoot growth: the length of 4-day-old etiolated hypocotyls decreased in the range from 46 % (oryzalin) to 27 % (EBL). Variants in which the orientation of cellulose in the cell walls became more chaotic (*pom2-4*, oryzalin, EBL) retained the ability to the acid growth, but to varying degrees. In the *pom2-4* mutant, the cell wall creep rate in the acidic buffer increased compared to the control (Col-0) (Fig. 1), whereas oryzalin (Fig. 2) and EBL decreased the rate of *in vitro* cell wall extension in Col-0 at pH 5.

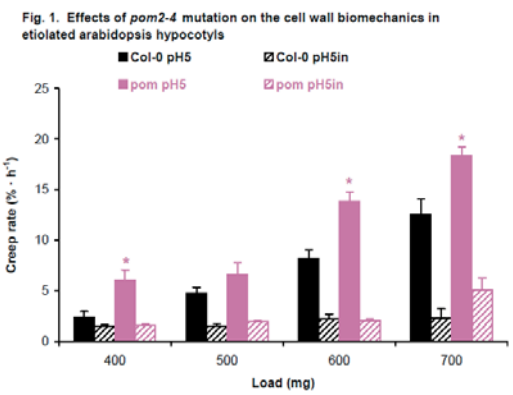


Fig. 1. Effects of *pom2-4* mutation on the cell wall biomechanics in etiolated arabidopsis hypocotyls

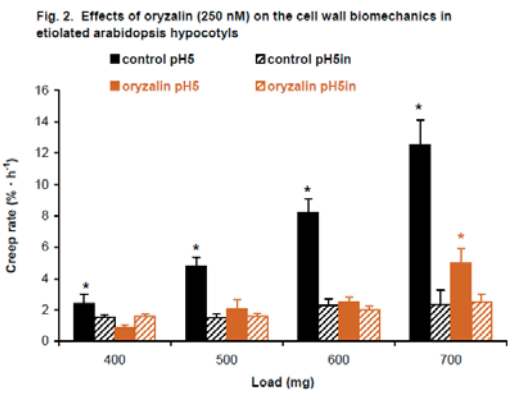


Fig. 2. Effects of oryzalin (250 nM) on the cell wall biomechanics in etiolated arabidopsis hypocotyls

Such a difference in the effects may be due to the unequal effect of these treatments on the hypocotyl cell wall cross-sectional area, since creep rate is proportional to the stress (force divided by the cross-sectional area of the sample) generated in the cell walls by a constant load. It was found that EBL significantly increased the cell wall cross-sectional area. Based on this, the cell wall stress in the presence of EBL was calculated and the dependence of the creep rate on the stress was determined. The character of this dependence demonstrates that the acidic buffer effect is opposite at low and high values of stress (Fig. 3A).

Fig. 3A. Effects of EBL (100 nM) on the biomechanics of native cell walls in etiolated Col-0 hypocotyls

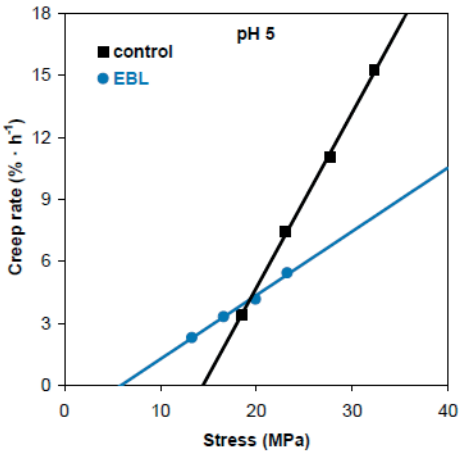
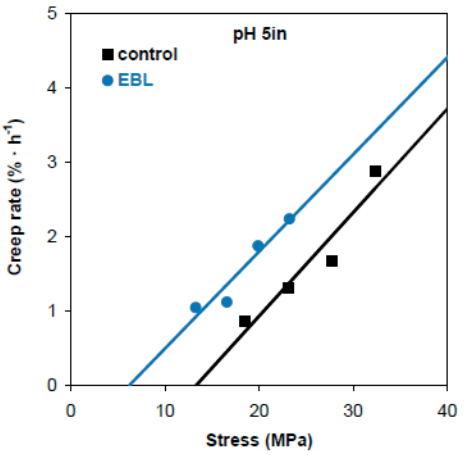


Fig. 3B. Effects of EBL (100 nM) on the biomechanics of heat-inactivated cell walls in etiolated Col-0 hypocotyls



In the mutant *ixr1-1*, characterized by reduced cellulose crystallinity, a powerful acidic buffer-induced creep of cell walls was detected, the rate of which was significantly higher than that of Col-0. Thus, a decrease in the crystallinity of cellulose microfibrils is not an obstacle to the acid growth of arabidopsis hypocotyls.

Under conditions of cellulose synthesis inhibition by isoxabene in Col-0 plants, the stimulation of cell wall creep with an acidic buffer was maintained, but was significantly lower than in Col-0 in the absence of the inhibitor (Fig. 1, 2). Curiously, in the presence of isoxabene, the creep rate at pH 5 changed very slightly with increasing load (or stress) in contrast to wild-type control plants, in which the acid-induced cell wall extension was significantly increased with increasing load (Fig. 1). This fact suggests a decrease in the activity of endogenous expansins under the influence of isoxabene. It is possible that under conditions of cellulose synthesis inhibition, the signaling cascades activated by cell wall integrity sensing are triggered [4], which suppress the expression of expansins.

Conclusion: Thus, in contrast to the literature data [2], it was shown in this study that the disorganization of cellulose under the influence of phytohormones, mutations and inhibitors does not completely eliminate, but modifies the acid growth of arabidopsis hypocotyls in different ways. The overall effect can be determined both by a change in the wall structure, rendering it more or less optimal for the work of expansins, and by a change in the level of expansin expression involving the systems of cell wall integrity sensing. The contribution of the latter mechanism could be clarified on the basis of transcriptomics data and subsequent bioinformatics analysis.

Funding: The study is supported by the RSF grant (No. 23-24-00379).

Список литературы/References

1. Arsuffi G., Braybrook S.A. Acid growth: an ongoing trip. *J Exp Bot*. 2018;69(2):137-146
2. Xin X., Lei L., Zheng Y., Zhang T., Pingali S.V., O'Neill H., Cosgrove D.J., Li S., Gu Y. Cellulose synthase interactive1- and microtubule-dependent cell wall architecture is required for acid growth in Arabidopsis hypocotyls. *J Exp Bot*. 2020;71(10):2982-2994
3. Somssich M., Vandenbussche F., Ivakov A., Funke N., Ruprecht C., Vissenberg K., Van Der Straeten D., Persson S., Suslov D. Brassinosteroids influence arabidopsis hypocotyl graviresponses through changes in mannans and cellulose. *Plant Cell Physiol*. 2021;62(4):678-692
4. Voxeur A., Höfte H. Cell wall integrity signaling in plants: “To grow or not to grow that's the question”. *Glycobiology*. 2016;26(9):950-960

doi 10.18699/bgrs2024-06-22

Множественность генов липоксигеназ в растительных геномах

Топоркова Я.*, Горина С., Смирнова Е., Ланцова Н., Ильина Т.,
Егорова А., Гречкин А.

Казанский институт биохимии и биофизики – обособленное структурное подразделение ФИЦ КазНЦ РАН, Казань, Россия

** yanchens@yandex.ru*

Ключевые слова: оксилипины; липоксигеназа; рекомбинантный фермент; субстратная специфичность; региоспецифичность

Мотивация и цель: Основным источником оксилипинов у растений является липоксигеназный каскад, который начинается с образования гидроперекисей при участии липоксигеназ (ЛОГ). Липоксигеназы представляют собой негемовые железо- или марганец-содержащие ферменты, которые катализируют регио- и стереоспецифическое перекисное окисление ненасыщенных жирных кислот, содержащих (1*Z*,4*Z*)-пентадиеновую систему, в соответствующие гидроперекиси. Липоксигеназы обнаружены у животных, растений, грибов и бактерий, водорослей и цианобактерий. Типичными субстратами растительных липоксигеназ являются линолевая и α-линоленовая кислоты. Продукты ЛОГ являются субстратами для ряда ферментов, среди которых редуктазы, пероксигеназы, а также цитохромы P450 семейства CYP74. При этом существует очевидный парадокс, поскольку ферментов дальнейшего превращения гидроперекисей, разнообразие которых достигает как минимум семи, в несколько раз меньше, чем самих липоксигеназ. Например, у огурца всего четыре фермента CYP74, а липоксигеназ – тридцать. При этом до недавнего времени все разнообразие липоксигеназ сводилось только к выделению 9- и 13-специфичных липоксигеназ, различающихся по атому углерода в составе линолевой или альфа-линоленовой кислот, который они окисляют. На объяснение этого парадокса и направлено данное исследование.

Методы и алгоритмы: Открытые рамки считывания целевых генов были клонированы в векторах для экспрессии белков системы pET. Рекомбинантные белки очищали методом анионообменной хроматографии. Продукты (гидроперекиси), очищенные с помощью ВЭЖХ, идентифицировали с помощью газовой хромато-масс-спектрометрии (ГХ-МС) и ЯМР, включая ^{1}H-ЯМР, 2D-COSY, HSQC и HMBC.

Результаты: Все исследования липоксигеназ, как правило, включают транскриптомные, биоинформатические, филогенетические исследования. Только небольшая часть работ включает исследования каталитических свойств, из них еще меньше проведено на рекомбинантных белках. В виде рекомбинантных белков получают, как правило, липоксигеназы невегетативных органов, таких как корни, клубни, семена.

Липоксигеназы семян сои являются классическими объектами. Липоксигеназа-1 описана как 13-специфичная, липоксигеназа-3 – как 9-специфичная. При этом липоксигеназа-2 аннотирована в базе данных NCBI как 9-специфичная, а все

исследования каталитических свойств описывают фермент как неспецифичный, окисляющий жирные кислоты по положению как 9, так и 13. Более того, в работах указаны разные соотношения образующихся продуктов. Однако ни в одной работе липоксигеназу-2 не получали в виде рекомбинантного белка. Мы получили рекомбинантный фермент. Было показано, что в результате инкубации фермента с линолевой и альфа-линоленовой кислотами образуются 13-гидроперекиси. Превращение других кислот проходит аналогично. Во всех реакциях образуются омега6-гидроперекиси.

В течение долгого времени исследователи не могли прийти к соглашению о том, каким образом входит субстрат в активный центр фермента. Изменение направленности (вход карбоксильным или метильным концом вперед) предполагается в качестве механизма регуляции для образования 9- или 13-гидроперекисей у неспецифичных липоксигеназ. Так же предполагалось для липоксигеназы-2 сои. Однако поскольку превращение всех жирных кислот протекает стереоспецифически, очевидно, что субстрат входит в активный центр липоксигеназы-2 сои только одним способом.

Автоокисление жирных кислот приводит к введению гидроперекисной группировки в углерод по разным положениям, в том числе по положению 16. При этом 16-гидроперекись альфа-линоленовой кислоты выявляется у разных растений и иногда в значительных количествах. Например, 16-ГПОТ была выявлена в разных органах огурца, при этом в цветках и кожуре огурца это была единственная гидроперекись. При этом наличие единственного стереоизомера указывает на ферментативное образование этой гидроперекиси.

Мы провели протеомный анализ разных органов огурца и выявили основные изоформы липоксигеназ. Одна из липоксигеназ присутствует как в листьях, так и в кожуре огурца. Мы клонировали открытую рамку считывания этого гена и получили рекомбинантный белок. Фермент проявляет предпочтение к альфа-линоленовой кислоте, стереоспецифически окисляя ее по углероду в положении 16, т. е. является омега3-специфичной липоксигеназой. Поскольку у линолевой кислоты нет двойной связи в положении омега3, превращение линолевой кислоты протекает неспецифически.

Выявление фермента, катализирующего образование омега3-гидроперекиси, имеет практическое значение, например, при производстве резольвинов, которые являются продуктами превращения 18-гидроперекиси С20 жирных кислот. Ранее эту гидроперекись получали при участии ферментов циклооксигеназы или некоторых цитохромов Р450.

Представленные выше результаты указывают как минимум на две причины, почему генов липоксигеназ в геномах растений в несколько раз больше, чем генов ферментов дальнейшего превращения гидроперекисей. Дополнительные причины этого феномена связаны с дупликацией генов липоксигеназ, а также с локализацией и кинетическими параметрами этих ферментов.

Выводы: Таким образом, выявлено несколько причин множественности генов липоксигеназ в геномах растений по сравнению с числом генов ферментов дальнейшего превращения гидроперекисей – продуктов липоксигеназной активности.

Финансирование: Биоинформатический и филогенетический анализ проводили при финансовой поддержке государственного задания Федерального исследовательского центра «КазНЦ РАН». Исследования рекомбинантных липоксигеназ поддержаны грантом Российского научного фонда 24-14-00418.

Multiplicity of lipoxygenase genes in plant genomes

Toporkova Y.*, Gorina S., Smirnova E., Lantsova N., Ilyina T.,
Egorova A., Grechkin A.
Kazan Institute of Biochemistry and Biophysics – a separate structural unit of the Federal Research Center Kazan Scientific Center of the Russian Academy of Sciences, Kazan, Russia
** yanchens@yandex.ru*

Key words: oxylipins; lipoxygenase; recombinant enzyme; substrate specificity; regiospecificity

Motivation and Aim: The main source of oxylipins in plants is the lipoxygenase cascade, which begins with the formation of hydroperoxides by lipoxygenases (LOXs). Lipoxygenases are non-heme iron- or manganese-containing enzymes that catalyze the regio- and stereospecific peroxidation of unsaturated fatty acids containing the (1*Z*,4*Z*)-pentadiene system resulting in the formation of the corresponding hydroperoxides. Lipoxygenases were found in animals, plants, fungi and bacteria, algae, and cyanobacteria. Typical substrates of plant lipoxygenases are linoleic and α-linolenic acids. LOX products, fatty acid hydroperoxides, are substrates for a number of enzymes, including reductases, peroxygenases, and cytochromes P450 of the CYP74 family. In this case, there is an obvious paradox, that there are several times fewer enzymes for the further conversion of hydroperoxides, the diversity of which reaches at least seven, than the lipoxygenases themselves. For example, cucumber has only four CYP74 enzymes and thirty lipoxygenases. Moreover, until recently, the entire variety of lipoxygenases was reduced only to 9- and 13-specific lipoxygenases, differing in the carbon atom within linoleic or alpha-linolenic acids, which they oxidize. This study is aimed at explaining this paradox.
Methods and Algorithms: The open reading frames of the target genes were cloned into pET system vectors for expressing proteins. Recombinant proteins were purified by anion exchange chromatography. Products (hydroperoxides) purified by HPLC were identified by gas chromatography-mass spectrometry (GC-MS) and NMR, including ^{1}H-NMR, 2D-COSY, HSQC, and HMBC.
Results: All studies of lipoxygenases, as a rule, include transcriptomic, bioinformatics, and phylogenetic studies. Only a small part of the work includes studies of catalytic properties, of which even less has been obtained as recombinant proteins. As a rule, lipoxygenases from non-vegetative organs, such as roots, tubers, and seeds, were obtained as recombinant proteins.
Soybean seed lipoxygenases are classical objects. Lipoxygenase-1 was described as 13-specific enzyme, lipoxygenase-3 was described as 9-specific one. At the same time, lipoxygenase-2 was annotated in the NCBI database as 9-specific enzyme, and all studies of catalytic properties described the enzyme as non-specific one, oxidizing fatty acids both at C9 and C13 positions. Moreover, different reports described different ratios of the products formed. However, no one report contained data obtained with lipoxygenase-2 as a recombinant protein. We have obtained a recombinant lipoxygenase-2. It was shown that as a result of incubation of the enzyme with linoleic and alpha-linolenic acids, 13-hydroperoxides were formed. The transformation of other acids was similar. All reactions produced omega-6 hydroperoxides.
For a long time, researchers could not agree on how the substrate enters the active site of the enzyme. Reversal of directionality (entry by carboxyl or methyl end first) was

proposed as a regulatory mechanism for the formation of 9- or 13-hydroperoxides in nonspecific lipoxygenases. The same mechanism was assumed for soybean lipoxygenase-2. However, since the conversion of all fatty acids occurs stereospecifically, it is obvious that the substrate enters the active site of soybean lipoxygenase-2 in only one way.

Autoxidation of fatty acids leads to the introduction of a hydroperoxide group into carbon at various positions, including position 16. In this case, 16-hydroperoxide of alpha-linolenic acid is detected in different plants and sometimes in significant quantities. For example, 16-HPOT was detected in various organs of cucumber. Moreover, in flowers and cucumber peel it was the only isomer. And the presence of a single stereoisomer indicates the enzymatic formation of this hydroperoxide.

We conducted a proteomic analysis of different cucumber organs and identified the main isoforms of lipoxygenases. One of the lipoxygenases is present in both the leaves and peel. We cloned the open reading frame of this gene and obtained a recombinant protein. The enzyme shows preference for alpha-linolenic acid, stereospecifically oxidizing it at carbon at position 16, so it is an omega3-specific lipoxygenase. Since linoleic acid does not have a double bond at the omega3 position, the conversion of linoleic acid occurs nonspecifically.

Identification of the enzyme that catalyzes the formation of omega3-hydroperoxide is of practical importance, for example, in the production of resolvins, which are products of the conversion of 18-hydroperoxide of C20 fatty acids. Previously, this hydroperoxide was obtained by cyclooxygenases or some P450 cytochromes.

The results presented above indicate at least two reasons why there are several times more genes for lipoxygenases in plant genomes than genes for enzymes for further conversion of hydroperoxides. Additional reasons for this phenomenon are associated with duplication of lipoxygenase genes, as well as the localization and kinetic parameters of these enzymes.

Conclusion: Thus, several reasons have been identified for the multiplicity of lipoxygenase genes in plant genomes in comparison with the number of genes for enzymes for the further conversion of hydroperoxides, products of lipoxygenase activity.

Funding: Bioinformatic and phylogenetic analyses were carried out with financial support from the state assignment of the Federal Research Center “KazSC of RAS”. Studies of the recombinant LOXs were supported by grant 24-14-00418 from the Russian Science Foundation.

doi 10.18699/bgrs2024-06-23

InterTransViewer: сравнительный анализ профилей дифференциальной экспрессии генов

Тяпкин А.[1], Омельянчук Н.[1, 2], Лавреха В.[1, 2], Землянская Е.[1, 2]*

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *ezemlyanskaya@bionet.nsc.ru*

Ключевые слова: транскриптом; метаанализ; ауксин; этилен; *Arabidopsis thaliana* L.

Мотивация и цель: Стремительный рост количества полногеномных экспериментов по изучению изменения экспрессии генов в различных условиях обусловил широкое распространение интегрированного анализа (метаанализа) транскриптомных данных. Интеграция данных может повысить точность статистических оценок, а также позволяет тестировать гипотезы, которые невозможно было проверить в отдельных исследованиях. Для повышения информативности такой интеграции необходимо оптимизировать подбор экспериментов. В данной работе мы предлагаем набор количественных показателей для всестороннего сравнительного описания транскриптомных данных, которые легко могут быть визуализированы и интерпретированы.

Методы и алгоритмы: Предлагаемые показатели включают в себя количество дифференциально экспрессирующихся генов (ДЭГ), долю специфических (уникальных) ДЭГ в каждом наборе данных, меру попарного сходства экспериментов по составу ДЭГ, статистическую оценку однородности профилей ДЭГ. Для автоматического вычисления и визуализации этих показателей мы разработали программу InterTransViewer.

Результаты: Программа InterTransViewer реализована в виде скрипта на языке R (v.4.1.2) и доступна по ссылке (https://github.com/al-t1/InterTransViewer0/). Мы применили InterTransViewer для сравнительного описания транскрипционных ответов на обработку фитогормонами у модельного растения *Arabidopsis thaliana* L. Проанализированы 23 ауксин-индуцированных и 16 этилен-индуцированных профилей дифференциальной экспрессии генов. Мы показали, что комплексное рассмотрение предлагаемых характеристик позволяет позиционировать эксперименты в контексте друг друга, оценивать тенденцию к их интеграции/сегрегации, генерировать гипотезы о влиянии нецелевых факторов на исследуемый транскрипционный ответ, выделять потенциально однородные группы экспериментов, статистически оценивать однородность этих групп профилей. Это помогает принять решение о целесообразности использования данных для метаанализа.

Выводы: В целом InterTransViewer позволяет эффективно формировать выборки экспериментов в зависимости от задачи и методов метаанализа.

Финансирование: работа выполнена при поддержке Российского научного фонда, грант № 20-14-00140.

InterTransViewer: a comparative analysis of differential gene expression profiles

Tyapkin A.[1], Omelyanchuk N.[1, 2], Lavrekha V.[1, 2], Zemlyanskaya E.[1, 2]*

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *ezemlyanskaya@bionet.nsc.ru*

Key words: transcriptome; meta-analysis; auxin; ethylene; *Arabidopsis thaliana* L.

Motivation and Aim: Meta-analysis of transcriptomic data from different experiments has become increasingly prevalent due to a significantly increasing number of genome-wide experiments investigating gene expression changes under various conditions. Data integration provides greater accuracy of statistic estimates and allows testing new hypotheses, which could not be validated in individual studies. To make data integration more informative, it is necessary to optimize the selection of experiments for the analysis. Here, we offer a set of quantitative indicators for a comprehensive comparative description of transcriptomic data, which can be easily visualized and interpreted.
Methods and Algorithms: The proposed indicators include the number of differentially expressed genes (DEGs), the proportion of experiment-specific (unique) DEGs in each data set, the measure of a pairwise similarity of experiments in DEG composition, and a statistic estimate of the homogeneity of DEG profiles. For automatic calculation and visualization of these indicators, we have developed an InterTransViewer program.
Results: InterTransViewer is implemented as an R script (v.4.1.2) and is available via the link (https://github.com/al-t1/InterTransViewer0/). We have used InterTransViewer to comparatively describe 23 auxin- and 16 ethylene-induced transcriptomes in *Arabidopsis thaliana* L. We have demonstrated that complex analysis of the proposed indicators allows ranking the experiments in the context of each other, assessing the tendency towards their integration or segregation, generating hypotheses about the influence of non-target factors on the transcriptional response, identifying potentially homogeneous groups of experiments, statistically estimating the homogeneity within the sets of DEG profiles. This helps to decide whether these data are appropriate for meta-analysis.
Conclusion: Overall, InterTransViewer makes it possible to efficiently select experiments for meta-analysis depending on its task and methods.
Funding: The study was supported by the RSF grant No. 20-14-00140.

doi 10.18699/bgrs2024-06-24

Оценка влияния гамма-облучения семян на развитие и активность ферментов антиоксидантной системы *Lupinus angustifolius* L.

Ханова А.*, Смирнова А., Блинова Я., Бондаренко Е.

ФГБУ «Всероссийский научно-исследовательский институт радиологии и агроэкологии Национального исследовательского центра «Курчатовский институт», Обнинск, Россия

** micenyk-anastasi@mail.ru*

Ключевые слова: *Lupinus angustifolius*; гамма-излучение; параметры прорастания; CAT; POX; APX

Мотивация и цель: Необходимым этапом для увеличения производства сельскохозяйственных кормовых культур с высоким содержанием легко усвояемого белка является расширение генетического разнообразия возделываемых зернобобовых культур, в частности холодостойкого и скороспелого люпина узколистного (*Lupinus angustifolius* L.). И радиационный мутагенез – это один из самых эффективных методов выведения новых сортов зернобобовых культур с различными ценными сельскохозяйственными признаками, поскольку частота возникновения спонтанных мутаций довольно низка [1]. В связи с этим для разработки протокола радиационного мутагенеза в отношении отечественных сортов люпина узколистного на примере сорта Белорозовый 144 была проведена серия радиобиологических экспериментов, финальная цель которых – определение оптимальных для индукции мутаций доз γ-излучения. Также целью работы являлась оценка влияния γ-излучения в дозах 25–1200 Гр на параметры прорастания, развитие и активность ферментов антиоксидантной системы проростков *L. angustifolius* для проведения ряда других фундаментальных радиобиологических исследований.

Методы и алгоритмы: Сухие семена отечественного сорта люпина узколистного Белорозовый 144 (любезно предоставленные ВНИИ люпина – филиал ФГБНУ «ФНЦ кормопроизводства и агроэкологии им. В.Р. Вильямса», г. Брянск) облучали в дозах 25, 50, 100, 200, 400, 800 и 1200 Гр при мощности дозы 90 Гр/ч на уникальной научной установке ГУР-120 (^{60}Co, ВНИИРАЭ) в трех повторностях по 30 семян в каждой. Культивирование контрольных и облученных семян люпина узколистного проводили в контролируемых условиях климатической камеры. Содержание ферментов антиоксидантной системы () анализировали при помощи спектрофотометра “NanoDrop OneC”. Статистическую обработку результатов осуществляли непараметрическими методами в средах программирования R (версия 4.2.3) и Python (версия 3.11).

Результаты: Взвешенный процент прорастания семян люпина узколистного статистически значимо уменьшился с увеличением дозы. Так, статистическая значимая разница выявлена при дозах 50 и 100 Гр и при дозах в диапазоне от 400 до 1200 Гр. С увеличением дозы также статистически значимо увеличилось среднее время прорастания семян по сравнению с контрольной группой, кроме группы, облученной в дозе 50 Гр. Скорректированный индекс энергии прорастания семян под влиянием гамма-излучения статистически значимо снижен

по сравнению с растениями в контрольной группе. Оценка фаз развития проростков люпина узколистного показала, что дозы 800 и 1200 Гр оказали негативное влияние как на скорость появления истинных листьев, так и на способность *L. angustifolius* их образовывать. В контроле и в группах, облученных в дозах 50, 100 и 200 Гр, истинные листья появились на 24 часа раньше (на пятые сутки прорастания), чем в группах, облученных в дозах 25 и 400 Гр (на шестые сутки прорастания). И на 48 часов раньше, чем в группах, облученных в дозах 800 и 1200 Гр (на седьмые сутки прорастания). Анализ показателей активности ферментов антиоксидантной системы в тканях первых истинных листьев люпина узколистного показал статистически значимое увеличение активности аскорбатпероксидазы (APX) в группах воздействия γ-излучения в дозах 800 и 1200 Гр по сравнению с контролем. Активность каталазы (CAT) также статистически значимо увеличилась в группах воздействия γ-излучения в дозе 50 Гр и в диапазоне доз от 200 до 1200 Гр. Что касается активности гваяколовой пероксидазы (POX), то показатель активность данного фермента статистически значимо повышен в группах, повергшихся γ-излучению в диапазоне доз от 50 до 1200 Гр. Также в результате эксперимента выявлена доза гамма-излучения, вызывающая 50 % уменьшение длины гипокотиля (RD50) для сорта Белорозовый 144 люпина узколистного. Величина RD50 составила 770 Гр.

Выводы: Таким образом, дозы выше 400 Гр для люпина узколистного при разработке протоколов радиационного мутагенеза будут расцениваться как ингибирующие. Для дальнейших радиобиологических исследований рекомендуется использовать дозы в диапазоне от 100 до 400 Гр.

Assessment of the impact of gamma irradiation of seeds on the development and activity of antioxidant enzyme systems in *Lupinus angustifolius* L.

Khanova A.*, Smirnova A., Blinova Ya., Bondarenko E.

Russian Institute of Radiology and Agroecology of National Research Centre "Kurchatov Institute", Obninsk, Russia

** micenyk-anastasi@mail.ru*

Key words: *Lupinus angustifolius*; gamma radiation; parameters of germination; CAT; POX; APX

Motivation and Aim: An essential step to increase the production of high-protein easily digestible forage crops is to expand the genetic diversity of cultivated grain legumes, in particular, the cold-resistant and fast-ripening narrow-leafed lupine (*Lupinus angustifolius* L.). Radiation mutagenesis is one of the most effective methods for breeding new varieties of grain legumes with various valuable agricultural traits, as the frequency of spontaneous mutations is quite low [1]. Therefore, a series of radiobiological experiments was conducted to develop a protocol for radiation mutagenesis regarding domestic varieties of narrow-leafed lupine, using the Belorozovyi 144 variety as an example. The ultimate goal was to determine the optimal doses of gamma radiation for inducing mutations. Additionally, the study aimed to assess the impact of gamma radiation doses ranging from 25 to 1200 Gy on germination

parameters, development, and antioxidant enzyme activity of *L. angustifolius* seedlings for conducting a series of other fundamental radiobiological research studies.
Methods and Algorithms: Dry seeds of the domestic variety of narrow-leafed lupine, Belorozovyi 144 (kindly provided by the All-Russian Research Institute of Lupine – branch of the Federal Scientific Center for Feed Production and Agroecology named after V.R. Williams, Bryansk) were irradiated at doses of 25, 50, 100, 200, 400, 800, and 1200 Gy with a dose rate of 90 Gy/h on the unique scientific setup GUR-120 (60Co, VNIIRAER) in three replications of 30 seeds each. Cultivation of control and irradiated narrow-leafed lupine seeds was carried out under controlled conditions in a climatic chamber. The content of antioxidant system enzymes was analyzed using a NanoDrop OneC spectrophotometer. Statistical analysis of the results was performed using non-parametric methods in programming environments R (version 4.2.3) and Python (version 3.11).
Results: The weighted percentage of germination of narrow-leafed lupine seeds significantly decreased with increasing dose. A statistically significant difference was found at doses of 50 and 100 Gy, as well as at doses ranging from 400 to 1200 Gy. Additionally, with increasing dose, the average germination time of seeds significantly increased compared to the control group, except for the group irradiated at 50 Gy. The corrected seedling energy index under the influence of gamma radiation was significantly reduced compared to plants in the control group. Assessment of the development phases of narrow-leafed lupine seedlings showed that doses of 800 and 1200 Gy had a negative impact on the speed of true leaf emergence and the ability of *L. angustifolius* to form them. In the control and groups irradiated at doses of 50, 100, and 200 Gy, true leaves appeared 24 hours earlier (on the fifth day of germination) than in groups irradiated at doses of 25 and 400 Gy (on the sixth day of germination). They also appeared 48 hours earlier than in groups irradiated at doses of 800 and 1200 Gy (on the seventh day of germination). Analysis of the activity of antioxidant system enzymes in the tissues of the first true leaves of narrow-leafed lupine showed a significant increase in ascorbate peroxidase (APX) activity in groups exposed to gamma radiation at doses of 800 and 1200 Gy compared to the control. Catalase (CAT) activity also significantly increased in groups exposed to gamma radiation at a dose of 50 Gy and in the dose range from 200 to 1200 Gy. Regarding guaiacol peroxidase (POX) activity, the activity of this enzyme was significantly increased in groups exposed to gamma radiation in the dose range from 50 to 1200 Gy. Additionally, the experiment identified the gamma radiation dose causing a 50 % reduction in hypocotyl length (RD50) for the Belorozovyi 144 variety of narrow-leafed lupine, with the RD50 value being 770 Gy.
Conclusion: Therefore, doses above 400 Gy for narrow-leafed lupine in the development of radiation mutagenesis protocols will be considered inhibitory. For further radiobiological studies, it is recommended to use doses in the range of 100 to 400 Gy.

Список литературы/References

1. Новик Н.В., Гераськин С.А., Якуб И.А. Влияние γ- облучения семян на внутрисортовую изменчивость количественных признаков люпина. Радиационная биология. *Радиоэкология.* 2022;62(6):620-628

doi 10.18699/bgrs2024-06-25

Границы цветков скользят внутри соцветий *Arabidopsis thaliana* с мутацией гена *APETALA1-1*

Харченко В.
ФГБОУ ВО «Луганский государственный аграрный университет им. К.Е. Ворошилова», Луганск, ЛНР, Россия
viktoriakharchenko@rambler.ru

Ключевые слова: *Arabidopsis*; *APETALA 1*; соцветие; цветок; гетерохрония

Мотивация и цель: Гомеотический ген *APETALA1* (*AP1*) кодирует предполагаемый фактор транскрипции, который действует локально, определяя идентичность цветочной меристемы и определяя развитие чашелистиков и лепестков. Он входит в состав хорошо известной ABC-модели генетического контроля цветочных органов. Эта модель основывается на анализе фенотипической изменчивости мутантов *Arabidopsis thaliana*, в числе которых мутант *apetala 1-1* (*ap1-1*) [1]. Несмотря на детальную изученность состава гена *APETALA1* и особенностей его экспрессии, все еще остаются вопросы, связанные с его функцией. Для полного представления о механизме генетического влияния целесообразно выяснить, какие морфогенетические изменения обуславливают наблюдаемый спектр изменчивости. У мутантов по гену *AP1* трансформируется не только околоцветник, а вся структура цветка и соцветия, которые действием локального гена объяснить невозможно. Эта работа фокусируется на анализе фенотипических эффектов у мутантов *ap1-1 A. thaliana* с целью уточнения функции гена *AP1*.
Методы и алгоритмы: Для исследований были использованы растения *A. thaliana* экотипа Landsberg. Семена растений были получены из Ноттингемского генетического центра хранения *Arabidopsis*. Мутация *ap1-1* является рецессивной, в каталоге NASC она имеет № 28. Линия *ap1-1* была получена в Университете Вагенингена, группой Maarten Koornneef (при воздействии этилметансульфонатом (EMS) на семена линии Landsberg *erecta* (L*er*). Согласно описанию фенотипа линии *ap1-1* в каталоге NASC, это сильный аллель, обуславливает гомеотическое превращение чашелистиков (первого оборота) в прицветники, в пазухах трансформированных чашелистиков образуется множество вторичных и третичных цветков; лепестки (второй оборот) обычно отсутствуют, но иногда заменяются тычинками или мозаикой, напоминающей лепестки-тычинки-прицветники. Фенотип изменяется в основании соцветия, и при низкой температуре цветки имеют более сильные преобразования, которые ослабевают в акропетальном направлении. Линия Landsberg *erecta* (L*er*) использовалась в качестве варианта контроля, в каталоге NASC она имеет № NW20. Фенотипическую изменчивость *A. thaliana* изучали в лаборатории светокультуры на кафедре биологии растений Луганского государственного аграрного университета и в лаборатории природной флоры Ботанического сада-института ДВО РАН (Владивосток). Стратификацию семян проводили при температуре +4 °С на протяжении 3 сут. Растения культивировали при температуре +20…+23 °С при круглосуточном освещении около 6000 люкс.

Фенотипы обеих линий были проанализированы по следующим критериям: число листьев прикорневой розетки (nrlMS), число листьев главного побега (nlMS), высота главного побега (hMS), общее число цветков на растении (nfP), число дней от посадки до начала цветения (PF), число дней от посадки до начала плодоношения (PFr), период вегетации (PV), число цветков в соцветии (единицы цветения) на главном побеге (nfUFMS), длина соцветия главного побега (LUFMS), плотность соцветия главного побега (pUFMS), длина бокового побега первого порядка, развивающегося из пазухи верхнего стеблевого листа (1.1Lls), длина соцветия на его верхушке (Luf1.ls) и число цветков на нем (nf1.1ls). Для анализа расположения цветочных зачатков на верхушке соцветия использовали микроскопы МБС-1, МБС-10 (с объективами ×8, ×20 и окулярами ×7, ×10, ×15), Биолам М-1 в ГОУ ВО ЛНР ЛГАУ, а также микроскоп Axioplan b Stemii в Ботаническом саду-институте ДВО РАН. Структура побегов была проанализирована на основании фрактального подхода. Для этого побег условно подразделяли на гомологичные трансформационные серии. Статистический анализ выполнен в программе STATISTICA при помощи тестов ANOVA. Для сравнения вариантов был использован дисперсионный анализ, результаты которого оценивали по F-критерию Фишера и силе влияния факторов, рассчитанных по методу Плохинского.

Результаты: В ходе проведенных исследований у растений мутантной линии *ap1-1 A. thaliana* был выявлен ряд фенотипических отличий от исходной линии L*er*. В частности, растения *ap1-1* имеют меньшую высоту и длину соцветий, но с большим числом голых цветков, у них развивается частично брактеозное соцветие тирс или плейотирс (множественный тирс), в отличие от эбрактеозной кисти, свойственной растениям исходной линии L*er*. Цветки L*er* дисимметричные, с двойным околоцветником, а у *ap1-1* цветки голые (или с фрагментами околоцветника), элементы в них могут располагаться по спирали, супротивно или мутовчато. При этом морфология андроцея варьирует по числу, расположению и длине тычинок. Новые цветки или оси соцветий могут формироваться на одном уровне с тычинками. Растения мутантной линии *ap1-1* зацветают раньше, но плодоношение и вегетация у них продолжается дольше. Кроме того, развитие прицветников у них смещено на более поздние стадии морфогенеза, чем у исходной линии. Следовательно, эта мутация является гетерохронной. У Brassicaceae развитие прицветников останавливается в зачаточном состоянии [2].

Структура цветка обычно очень консервативна, но у *A. thaliana* линии *ap1-1* она сильно варьирует, а спектр изменений включает структуру соцветия. Цветки имеют мутовчатый филлотаксис, который приводит к образованию устойчивой системы с повышенной синорганизацией. Поэтому усиление давления прицветников на «меристему цветка» может потревожить филлотаксический рисунок цветка и стать причиной хаотического заложения его элементов [3]. В ходе морфогенеза из-за гетерохронии действие механических сил на «меристему цветков» может усиливаться или ослабевать и таким образом обуславливать поливариантность структуры цветков и соцветий [4].

Выводы: Формирование прицветников в соцветиях мутантов *ap1-1 A. thaliana* свидетельствует о том, что их развитие смещается с эмбриональной стадии на более поздние стадии морфогенеза. У *ap1-1* развитие прицветников смещается на более поздние стадии морфогенеза. Это инициирует смещение во времени развития органов цветка. Следовательно, мутация *ap1-1 A. thaliana* является

гетерохронной, а цветки и прицветники в соцветии представляют единую трансформационную серию. По-видимому, одной из функций гена *APETALA 1* является регуляция морфогенеза элементов внутри соцветия. Учитывая спектр изменений, обусловленных мутацией *ap1-1* можно предположить**,** что ген *APETELA 1* обуславливает сепарацию элементов внутри цветка и внутри соцветия.

Floral boundaries are sliding inside of inflorescences *Arabidopsis thaliana* with mutation of the gene *APETALA1-1*

Kharchenko V.*

FSBEI HE «Lugansk state agrarian university named after K.E. Voroshilov», Lugansk, LNR, Russia

** viktoriakharchenko@rambler.ru*

Key words: *Arabidopsis*; *APETALA 1*; inflorescence; flower, heterochrony

Motivation and Aim: The gene *APETALA1* (*AP1*) encodes a putative transcription factor that acts locally to determine the identity of the floral meristem and determines sepal and petal development. It is part of the well-known ABC model of genetic control of floral organs, which is based on an analysis of the phenotypic variability of *Arabidopsis thaliana* mutants, including *apetala 1-1* (*ap1-1*) [1]. Despite the detailed study of *AP1* gene composition and its expression patterns, there are still questions related to its function. For a complete understanding of the mechanism of genetic influence, it is useful to find out what morphogenetic changes are responsible for the observed spectrum of variability? In *AP1* mutants, not only the perianth is transformed, but the whole structure of the flower and inflorescence, which cannot be explained by the action of a local gene. This work focuses on analysing phenotypic effects in *ap1-1* mutants of *A. thaliana* in order to clarify the function of the *AP1* gene.

Methods and Algorithms: Plants of *A. thaliana* ecotype Landsberg were used for the studies. Plant seeds were obtained from the Nottingham Arabidopsis Genetic Storage Center. The *ap1-1* mutation is recessive and has No.28 in the NASC catalog. The *ap1-1* line was obtained at Wageningen University, Maarten Koornneef's group (by exposing seeds of the line Landsberg *erecta* (L*er*) to ethyl methanesulfonate (EMS). According to the description of the phenotype of line *ap1-1* in the NASC catalog, this strong allele causes homeotic transformation of sepals (first turn) into bracts, many secondary and tertiary flowers are formed in the axils of transformed sepals; petals (second turn) are usually absent, but are sometimes replaced by stamens or by a mosaic resembling petals-petals-stamens-bracts. The phenotype changes at the base of the inflorescence and at low temperatures, flowers have stronger transformations that weaken in the acropetal direction. Phenotypic variability of *A. thaliana* was studied in the laboratory of light culture at the Department of Plant Biology of Lugansk State Agrarian University and in the laboratory of natural flora of the Botanical Garden of the Institute of the Far East Branch of the Russian Academy of Sciences (Vladivostok). Seed stratification was carried out at +4 °C for 3 days. Plants were cultivated at +20…+23 °C, under 24-hour illumination of about 6000 lux. The phenotypes of both lines were analysed according to the following criteria: number of root rosette leaves (nrlMS), number of main shoot leaves (nlMS), main shoot height (hMS), the total number of flowers per plant (nfP), number of days from planting to flowering initiation (PF), number of days from planting to fruiting initiation (PFr), growing season (PV), number of flowers per inflorescence

(flowering units) on the main shoot (nfUFMS), main shoot inflorescence length (LUFMS), main shoot inflorescence density (pUFMS), length of first-order lateral shoot developing from the axil of the upper stem leaf (1. 1Lls), the length of the inflorescence at its apex (Luf1.1ls) and the number of flowers on it (nf1.1ls). Microscopes MBS-1, MBS-10 (with lenses ×8, ×20 and eyepieces ×7, ×10, ×15), Biolam M-1 at the State Educational Institution of Higher Professional Education LNR LSAU, and Axioplan b Stemii microscope at the Botanical Garden-Institute of the Far East Branch of the Russian Academy of Sciences were used to analyze the location of floral rudiments at the inflorescence apex. The shoot structure was analyzed based on the fractal approach. For this purpose, shoots were conditionally subdivided into homologous transformation series. Statistical analysis was performed in STATISTICA program using ANOVA tests. The analysis of variance was used to compare the variants, the results of which were evaluated by Fisher's F-criterion and the strength of the influence of factors calculated by the Plokhinsky method.

Results: In the course of these studies, plants of the *ap1-1* mutant line of *A. thaliana* showed a number of phenotypic differences from the original line L*er*. In particular, *ap1-1* plants have smaller inflorescence height and length, but with a greater number of glabrous flowers, and they develop a partially bracteose inflorescence thyrsus or pleiotyrsus (multiple thyrsus), in contrast to the ebracteose tassel characteristic of plants of the original L*er* lineage. Flowers of L*er* are disymmetrical, with double perianth, while in *ap1-1* flowers are glabrous (or with fragments of perianth), the elements in them can be arranged spirally, supratrophically, or whorled. New flowers or inflorescence axes may form at the same level as the stamens. Plants of the *ap1-1* mutant line flower earlier, but their fruiting and vegetation last longer; in addition, the development of bracts is shifted to later stages of morphogenesis than in the original line. This initiates a time shift in the development of flower organs. Consequently, this mutation is heterochronic. In *ap1-1*, bract development is shifted to later stages of morphogenesis. In Brassicaceae, the development of bracts stops in infancy. Flower structure is usually very conservative, but in *A. thaliana* line *ap1-1* it is highly variable, and the spectrum of variation includes inflorescence structure [2]. Flower structure is usually very conservative, but in *ap1-1* it is highly variable, and the spectrum of variation includes inflorescence structure. Flowers have whorled phyllotaxis, which results in a stable system with increased syn-organization. Therefore, increased pressure of bracts on the "flower meristem" may disturb the phyllotactic pattern of the flower and cause chaotic lying of its elements [3]. During morphogenesis, due to heterochrony, the action of mechanical forces on the "flower meristem" can strengthen or weaken and thus determine the polyvariance structure of flowers and inflorescences [4].

Conclusion: The formation of bracts in inflorescences of *ap1-1* mutants of *A. thaliana* indicates that their development shifts from the embryonic stage to later stages of morphogenesis. In ap1-1, bract development is shifted to later stages of morphogenesis. This initiates a shift in the timing of flower organ development. Consequently, the *ap1-1* mutation of *A. thaliana* is heterochronic, and the flowers and bracts in the inflorescence represent a single transformation series. Apparently, one of the functions of the *APETALA 1* gene is the regulation of morphogenesis of elements within the inflorescence. Given the spectrum of changes caused by the *ap1-1* mutation, it can be assumed that the *APETALA 1* gene is responsible for the separation of elements within the flower and within the inflorescence.

Список литературы/References

1. Bowman J.L., Smyth D.R., Meyerowitz E.M. The ABC model of flower development: then and now. *Development.* 2012;139(22):4095-4098. doi 10.1242/dev.083972
2. Kwiatkowska D. Flowering and apical meristem growth dynamics. *J Exp Bot.* 2008;59:187-201
3. Endress P.K. Development and evolution of extreme synorganization in angiosperm flowers and diversity: a comparison of Apo ynaceae and Orchidaceae. *Ann Bot.* 2016;117:749-767
4. Ronse De Craene L. The interaction between heterochrony and mechanical forces as main driver of floral evolution. *J Plant Res.* 2024. doi 10.1007/s10265-024-01526-3

doi 10.18699/bgrs2024-06-26

Новая методика количественного описания процесса обновления клеток в корневом чехлике у *Arabidopsis thaliana* L.

Черенко В.[1, 2]*, Землянская Е.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *cherenkova@bionet.nsc.ru*

Ключевые слова: корневой чехлик; *Arabidopsis thaliana*; микроскопия; фенотипирование; ауксин

Мотивация и цель: Понимание того, с помощью каких механизмов определяется размер органов у живых организмов, является одной из важнейших задач биологии развития. Одна из стратегий такой регуляции – координация пространственно разобщенных процессов деления и удаления клеток. У растений эта стратегия реализуется в корневом чехлике (КЧ), который располагается на кончике корня, защищая от повреждений нишу стволовых клеток. КЧ непрерывно пополняется новыми клетками в результате деления инициалей, при этом его размер и форма остаются постоянными [1]. Поддержание размера и структуры этого органа крайне важно для правильного функционирования корня и происходит благодаря синхронизации делений стволовых клеток колумеллы (центральной части корневого чехлика) и удаления внешнего слоя клеток корневого чехлика [2]. Однако механизмы, координирующие процессы деления и слущивания клеток в корневом чехлике, мало изучены. Для эффективного исследования механизмов, координирующих эти процессы, необходима простая скрининговая методика, позволяющая количественно оценивать параметры клеточной динамики в КЧ. Разработка такой методики применительно к модельному растению *Arabidopsis thaliana* является целью данной работы.

Методы и алгоритмы: Методика предназначается для относительной оценки изменения скоростей деления и удаления клеток в корневом чехлике при изменении условий. Для этого использовали легко детектируемые структурные параметры КЧ: количество отделившихся клеточных слоев (у *A. thaliana* клетки КЧ слущиваются единым слоем, а не по отдельности, что позволяет легко оценить этот параметр), количество клеточных слоев без признаков их отделения от КЧ и общее количество клеточных слоев. В основе методики – два этапа микроскопии для каждого исследуемого корня: микроскопия в проходящем свете без переноса с твердой питательной среды с целью оценки количества отделившихся клеточных слоев КЧ; конфокальная микроскопия с целью оценки количества клеточных слоев без признаков их отделения от КЧ.

Результаты: Считается, что ключевым координатором деления стволовых клеток и слущивания клеток дистального слоя корневого чехлика является градиент фитогормона ауксина с максимумом в стволовых клетках и минимумом в клетках отделяющегося слоя [3]. Мы провели апробацию разработанной методики для оценки влияния ауксина (индол-3-уксусная кислота) на клеточную динамику в КЧ *A. thaliana.*

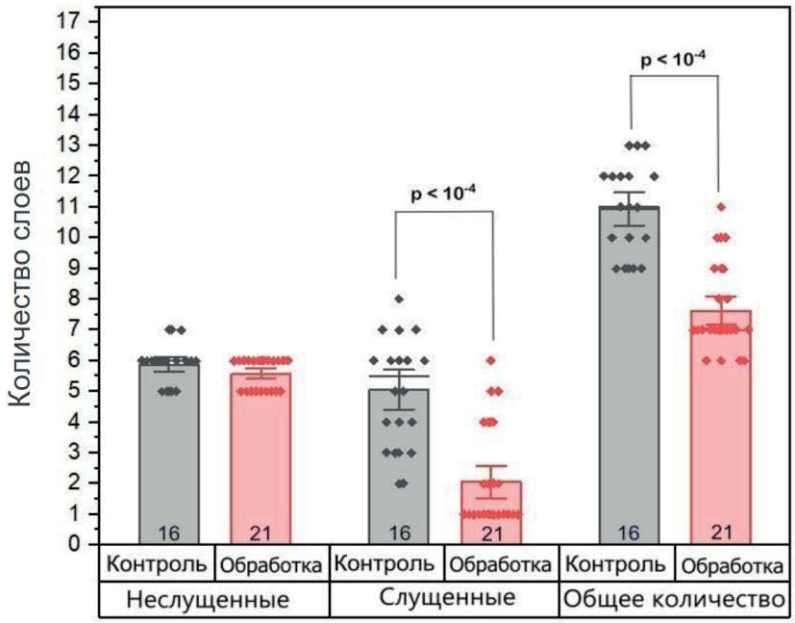


Рис. 1. Количество слоев в корневом чехлике после обработки ауксином в концентрации 0.1 мкМ
На гистограмме отображено среднее значение ± стандартная ошибка. Медиана отображена горизонтальной линией. Размер выборки указан внутри столбцов

Нам удалось зафиксировать изменения структурных параметров КЧ (рис. 1), а именно снижение количества слущенных слоев КЧ и общего количества слоев; количество слоев без признаков отделения от КЧ осталось неизменным. Таким образом, показано, что избыток ауксина сбалансированно снижает скорость деления инициалей КЧ и скорость слущивания наружного клеточного слоя.
Выводы: Предлагаемая нами методика может быть успешно использована для количественной оценки параметров клеточной динамики в КЧ на фоне мутаций в потенциальных регуляторах.
Финансирование: Исследование поддержано грантом РНФ № 20-14-00140.

New technique for the quantitative description of root cap renewal in *Arabidopsis thaliana* L.

Cherenko V.[1, 2]*, Zemlyanskaya E.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *cherenkova@bionet.nsc.ru*

Key words: root cap; *Arabidopsis thaliana*; microscopy; phenotyping; auxin

Motivation and Aim: Understanding the mechanisms, which determine organ size in living organisms is one of the most important tasks of developmental biology. One of the strategies of this regulation is to coordinate spatially separated cell division and cell removal. In plants, this strategy is implemented in the root cap (RC), which is located at

the root tip, protecting the stem cell niche from damage. The RC is continuously replenished with new cells due to the division of the initials, and its size and shape remain constant [1]. The maintenance of the size and structure of this organ is essential for the proper functioning of the root and occurs due to the synchronization of the division of the stem cells in the columella (central part of the root cap) and the removal of the outer layer of the root cap cells [2]. However, the mechanisms that coordinate cell division and removal in the root cap are poorly understood. In order to effectively study the mechanisms coordinating these processes, a simple screening technique is needed, which allows quantification of the parameters of cell dynamics in the RC. The development of such a technique for the model plant *Arabidopsis thaliana* is the goal of this work.

Methods and Algorithms: The technique is designed for the relative assessment of changes in cell division and removal rates in the root cap under changing conditions. To achieve this, easily detectable structural parameters of the root cap were utilized: detached cell layers (in *A. thaliana*, RC cells slough off as a single layer rather than individual cells, which allows for easy estimation of this parameter), the number of cell layers without signs of detachment from the RC, and the total number of cell layers. The methodology is based on two microscopy stages for each examined root: bright-field microscopy without transfer from solid nutrient medium to assess the number of detached cell layers in the RC; confocal microscopy to assess the number of cell layers without signs of detachment from the RC.

Results: The gradient of the phytohormone auxin with a maximum in the stem cells and a minimum in the cells of the separating layer is thought to be a key coordinator of stem cell division and the detachment of cells from the root cap surface [3]. We used the new technique to assess the influence of auxin (indole-3-acetic acid) on cell dynamics in the root cap of *A. thaliana*.

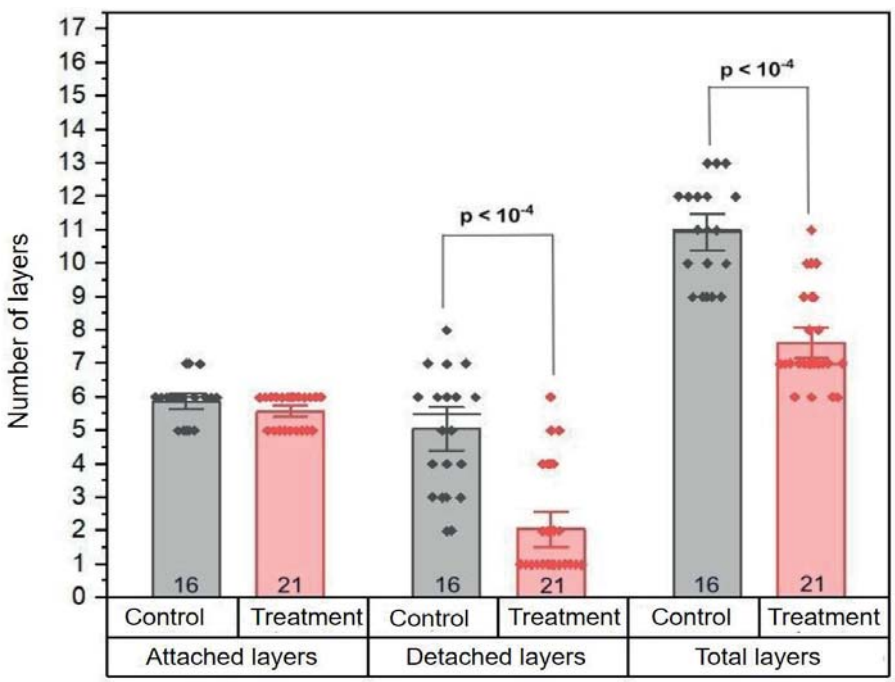


Fig. 1. Number of layers in the root cap after auxin treatment at a concentration of 0.1 μM. The chart shows the average value ± the standard error. Median is designated as a horizontal line. Number of samples is shown in the bars

We observed changes in the structural parameters of the root cap (Fig. 1), namely: a decrease in the number of detached layers of the root cap and the total number of layers, while the number of layers without signs of detachment from the RC remained unchanged. Thus, it was demonstrated that an excess of auxin evenly reduced the division rate of root cap initials and the rate of sloughing of the outer cell layer.

Conclusion: The proposed methodology can be successfully used to estimate changes in the parameters of cell dynamics in RC in mutant background with the deficiency in the potential regulators.

Funding: The study was supported by the RSF grant No. 20-14-00140.

Список литературы/References

1. Kumpf R.P., Nowack M.K. The root cap: a short story of life and death. *J Exp Bot*. 2015;66(19):5651-5662. doi 10.1093/JXB/ERV295
2. Fendrych M., Hautegem T.V. et al. Programmed cell death controlled by ANAC033/SOMBRERO determines root cap organ size in Arabidopsis. *Curr Biol*. 2014;24(9):931. doi 10.1016/J.CUB.2014.03.025
3. Dubreuil C., Jin X., Grönlund A., Fischer U. A local auxin gradient regulates root cap self-renewal and size homeostasis. *Curr Biol*. 2018;28(16):2581-2587. doi 10.1016/J.CUB.2018.05.090

doi 10.18699/bgrs2024-06-27

Genome-wide association study reveals novel loci modulating pre-harvest sprouting and grain color in red-grained winter wheat *T. aestivum* L.

Afonnikova S.[1, 2, 3]*, Kiseleva A.[1, 3], Fedyaeva A.[1], Komyshev E.[1, 3], Koval V.[1, 3], Afonnikov D.[1, 2, 3], Salina E.[1, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Faculty of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

[3] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *svetaafonnikova@gmail.com*

Key words: red-grained wheat; seed coat color; pre-harvest sprouting; association analysis; image analysis

Motivation and Aim: The association between pre-harvest sprouting (PHS) and grain color has long been recognized. Red-grained wheats generally are more PHS resistant compared to white-grained wheat. Nevertheless, red-grained wheat vary greatly in PHS resistance [1, 2]. The search for genes controlling PHS in a color-independent manner is of special interest. For instance, developing PHS-resistant cultivars in white-grained wheats is advantageous due to their lighter flour color and higher flour yield [3]. Research based on analysis of population with identical color allows to identify the novel gene loci involved in PHS control [4, 5]. In present work, we conducted a genome-wide association study on a panel of red-grained wheat varieties, aimed at uncovering genes that modulate PHS resistance and red color components of seed coat using digital image processing.

Methods and Algorithms: Statistical analysis of pre-harvest sprouting and grain color traits (brightness, lightness, redness, yellowness and blueness), genome-wide association analysis testing MLM, FarmCPU and BLINK models, gene prioritization using gene expression data from expVIP, functional annotation using InterPro and DAVID resources, and the literature.

Results: We used a panel of red-grained winter wheat accessions exhibiting variations from high to low in resistance to PHS as determined by the analysis of germination index at the late milk/hard dough (GI_milk) and mature grain stages (GI_mat) and the falling number at mature grain stage (FN). The heritability of all PHS-related traits was estimated above 50 %. The variability of grain color characteristics was detected by a digital phenotyping, their heritability varied from low (below 20 %) up to 50 %. The correlation analysis showed a weak, predominantly statistically insignificant relationships between grain color and PHS traits.

As a result of genome-wide association analysis we identified 12 markers associated with PHS traits at $p < 10^{-4}$ on chromosomes 1A, 1B, 1D, 3A, 4A, 5B, 6B, and 7D. Particularly, SNPs on chromosome 1B and chromosome 7D each of them explained a up to 25.3 % of the phenotypic variance. Eight loci, containing the detected markers, were considered novel. We observed 26 markers associated with seed color traits at $p < 10^{-4}$. They were located on the chromosomes 1A, 1B, 1D, 2A, 2B, 3B, 5A, 6A and 7D. Markers on chromosome 1D explained up to 50 % of brightness, redness and blueness. Some SNPs were associated with multiple traits for color characteristics.

Interestingly, loci associated with color traits did not overlap with loci associated with PHS-related traits.

As a result of gene prioritization, we detected 12 candidate genes for PHS-related traits. The most promising gene was *TraesCS6B02G147900*, encoding a protein involved in aleurone layer morphogenesis. Out of 14 genes prioritized for the grain color two genes, *TraesCS1D03G0758600* and *TraesCS7B03G1296800*, were considered as top candidates. The former is located on chromosome 1D and encodes a flavonol-synthase which synthesizes flavonols in the flavonoid biosynthesis pathway. The second gene is located on chromosome 7B and encodes a phytoene synthase which participates in one of the initial steps in carotenoid biosynthesis.

Conclusion: In our study of a population of red-grained common wheat varieties, we found that variation in coloration among these varieties does not correlate with resistance to PHS. Thus, when analyzing populations of wheat with red grain, we exclude grain color as a factor from our analysis. This exclusion allowed us to detect novel loci in the genome associated with PHS and grain color and suggest significant SNPs that could potentially enhance wheat varieties during the breeding.

Through the analysis of expression data and gene annotation within the revealed loci, we identified candidate genes conferring PHS resistance, in addition to the known genes. Our analysis suggests that the structure and composition of the aleurone layer may play a role in this process. Furthermore, we propose that variation in grain color characteristics in red-grained wheat, such as brightness, redness and blueness, is regulated by the balance between the synthesis of uncolored flavonols and colored anthocyanins, regulated by the activity of flavonol synthase. Another speculated mechanism underlying grain color is the carotenoid synthesis and accumulation pathway controlled by phytoene synthase.

Funding: The study is supported by the Russian Science Foundation project No. 22-74-00122 (grain color trait phenotyping and analysis) and project No. 21-76-30003 (genetic data analysis).

References

1. Groos C., Gay G., Perretant M.-R., Gervais L., Bernard M., Dedryver F., Charmet G. Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white×red grain bread-wheat cross. *Theor Appl Genet.* 2002;104(1):39-47. doi 10.1007/s001220200004
2. Albrecht T., Oberforster M., Kempf H. et al. Genome-wide association mapping of preharvest sprouting resistance in a diversity panel of european winter wheats. *J Appl Genet.* 2015;56(3):277-285. doi 10.1007/s13353-015-0286-5
3. Fakthongphan J., Bai G., St. Amand P., Graybosch R.A., Baenziger P.S. Identification of markers linked to genes for sprouting tolerance (independent of grain color) in hard white winter wheat (HWWW). *Theor Appl Genet*. 2016;129(2):419-430. doi 10.1007/s00122-015-2636-4
4. Bi H.H., Sun Y.W., Xiao Y.G., Xia L.Q. Characterization of DFR Allelic variations and their associations with pre-harvest sprouting resistance in a set of red-grained chinese wheat germplasm. *Euphytica.* 2014;195(2):197-207. doi 10.1007/s10681-013-0986-z
5. Liu S., Cai S., Graybosch R., Chen C., Bai G. Quantitative trait loci for resistance to pre-harvest sprouting in us hard white winter wheat Rio Blanco. *Theor Appl Genet.* 2008;117(5):691-699. doi 10.1007/s00122-008-0810-7

doi 10.18699/bgrs2024-06-28

Development of an algorithm for separating densely contiguous wheat grains in 2D images

Avzalov D.R., Komyshev E.G., Afonnikov D.A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

Key words: wheat grain; contour analysis; image recognition; computer vision

Motivation and Aim: Counting grains manually is a labor-intensive and time-consuming task. Development of approaches based on image analysis by computer vision methods is relevant nowadays due to low cost and ease of use. Grains in such images are located close to each other, which complicates the process of their recognition. In addition, the characteristics of individual grains, such as length, width, volume, shape, color, and texture, are important in modern breeding and genetic experiments for the measurement of which grain boundaries must be precisely defined.
Methods and Algorithms: The algorithm consists of 2 main steps: preprocessing the image to obtain contours of grain groups and analyzing the obtained contours to separate contiguous seeds. The preprocessing consists of reducing the resolution of the image, removing noise using the Gaussian filtering method, and obtaining a binary mask using the Mean shift clustering algorithm and the OTSU algorithm. A combination of the corner point detection algorithm from [1, 2] and the elliptical contour analysis algorithm was used to analyze the obtained contours.
Results: The metric CR (correct ratio) [2] was used to evaluate the performance of the algorithm. A comparison was made with other commonly used classical computer vision approaches for object segmentation: the watershed algorithm and the erosion algorithm. The average value of the metric was 96 %, which is a higher result compared to other classical computer vision approaches.
Conclusion: The approach based on classic computer vision algorithms does not require image markup and allows counting the number and finding the contours of grains in loose arrangements. The algorithm using an approximation of the contour by ellipses significantly increases the accuracy of counting the number of grains using the corner point search algorithm.
Funding: The work was supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, in agreement with the Ministry of Education and Science of the Russian Federation, No. 075-15-2019-1662.

References

1. Tan S. et al. Segmentation and counting algorithm for touching hybrid rice grains. *Comp Electron Agriculture*. 2019;162:493-504
2. Zhang J. et al. Research on a rapid identification method for counting universal grain crops. *PloS One*. 2022;17(9): e0273785

doi 10.18699/bgrs2024-06-29

Using YOLO neural network for segmentation of contiguous wheat grains

Avzalov D.R., Komyshev E.G., Afonnikov D.A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

Key words: wheat grain; computer vision; instance segmentation; yolo

Motivation and Aim: The weight of a thousand grains is one of the most important yield indicators of agricultural cereal crops. Its calculation requires measuring the weight and counting the number of seeds. Counting grains manually is a labor-intensive and time-consuming task. Development of approaches based on image analysis by computer vision methods is relevant at the moment because of low cost and ease of use. Grains in such images are located close to each other, which complicates the process of their recognition. In addition, the characteristics of individual grains, such as length, width, volume, shape, color, and texture, are important in modern breeding and genetic experiments. To measure them in the image, it is important to accurately define grain boundaries.
Methods and Algorithms: The study involved the selection of five genotypes from the ITMI collection, with attention to cover genotypes having different grain colors. The neural network architecture from the Yolo [1] family of the eighth version was chosen to solve the instance segmentation problem. The model was trained for detection and segmentation of individual grains. The training sample size was increased using the image augmentation method. The training and model running parameters such as number of epochs, batch size, optimizer? learning rate, IoU, confidence were selected.
Results: The quality of the model performance on the test sample was evaluated using metrics such as CR (correct ratio) [2] and mAP (mean average precision). CR was used to assess the quality of counting and was equal to 98.3 %, which is a higher result compared to classical computer vision methods often used for solving similar problems. The mAP metrics were computed to evaluate both detection and segmentation quality. The metrics for detection have higher values compared to the metrics for segmentation.
Conclusion: Neural network allows to find contours in images with dense arrangement of grains, in which application of methods of classical computer vision will not give high results. The result of segmenting individual grains can be further processed using neural network models per-designated to solve the semantic segmentation problem or using classical computer vision techniques.
Funding: The work was supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, in agreement with the Ministry of Education and Science of the Russian Federation, No. 075-15-2019-1662.

References

1. Terven J., Córdova-Esparza D.M., Romero-González J.A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Mach Learn Knowl Extr*. 2023;5(4):1680-1716
2. Zhang J. et al. Research on a rapid identification method for counting universal grain crops. *PloS One*. 2022;17(9):e0273785

doi 10.18699/bgrs2024-06-30

DyCeModel: a tool for 1D simulation for plant hormones control of tissue patterning using shoot and root apical meristems of *Arabidopsis thaliana* as examples

Azarova D.S.[1], Zemlyanskaya E.V.[1, 2], Lavrekha V.V.[1]*

[1] *Department of Systems Biology, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

* *vvl@bionet.nsc.ru*

Key words: computer modelling; developmental trajectory; input data; genetic algorithm; plant hormones

Motivation and Aim: The question of the mechanisms of growth and development is very challenging. To do this, it is necessary to analyze changes in the regulators distribution throughout the tissue in time and space and take into account their influence on cellular dynamics within the tissue. The main regulators of cellular dynamics in tissues are plant hormones, which form gradients and maxima to control molecular processes depending on concentration. Previously mathematical modeling was successfully used to study the effect of the plant hormones concentration distribution on the plant stem cell niches functioning in 1D and 2D models, in which cell division, growth and differentiation were under the control of signaling molecules [1–5]. Nevertheless, these models remain within a specific meristem and are not suited to other plant stem cell niches. A common description of the main complex of processes associated with the redistribution of hormone concentrations and cell responses to them can serve as the basis for studying the general and specific characteristics of meristems in various plants. Therefore, one of the actual tasks is to create software that builds computer models of various plant organs using uniformly described processes and provides automatic adjustment of parameters. We have created the DyCeModel pipeline for one-dimensional tissue modeling with a dynamic cellular ensemble, where changes in the concentration of hormones (or other active substances) in cells are described by ordinary differential equations (ODEs).
Methods and Algorithms: The DyCeModel pipeline allows you to create a dynamic one-dimensional structure of cells, integrate it into a mathematical model in the form of an ODE, and conduct a computational experiment. It contains five script files (.m files) that run in the MATLAB software environment. The substance_eq.m file contains an ODE system for describing the synthesis, degradation, passive and active transport of substances of interest. By default, DyCeModel provides examples of functions describing these processes for two substances in accordance with Michaelis-Menten kinetics and the Hill generalized function method [6], Fick's diffusion law and the mass action law (to describe active transport). Alternatively, users can create their own functions instead of the default ones. The parameters_fitting.m file describes the implementation of a genetic algorithm for assessing the similarity of the simulated distribution to experimental data. The model_parameters.m file contains all the model parameter default values for the ODE system and describes the model configuration (substance inflows). The grow_eq.m file describes the cell growth function, and tool_1d_model.m file provides the modeling scenario. It is important to note that there are two different strategies for using DyCeModel: using the parameters_fitting.m file or

not using it. In the latter case, the user should define all parameters in the model_parameters.m file.

Results: The DyCeModel tool builds mathematical models of hormone distribution based on the processes of their synthesis, degradation, diffusion and active transport in a dynamically developing cellular ensemble. As DyCeModel allows to model different types of plant meristems based on the same ODEs with a set of certain input characteristics for different meristems, we demonstrated the effectiveness of the tool by modeling the influence of main plant hormones auxin and cytokinin on tissue patterning in two types of *Arabidopsis thaliana* stem cell niches: root and shoot apical meristems. Using DyCeModel, we built the models of the root and shoot apical meristems by automatically selecting model parameters that matched the experimental data. For automatically parameters fitting we used the following restrictions on parameter values: approximately equal values of diffusion parameters for auxin and cytokinin and a low level of auxin synthesis. Obtained hormone distribution profiles in the models were qualitatively consistent with published experimental data. In result, proliferative activity zones were identified and the number of cells within those zones were maintained at a certain level throughout the entire time of model calculations. At the same time, the identified parameters of passive transport, degradation and growth were the same for the shoot meristem and root meristem models, and the parameters determining cell division remained similar.

Conclusion: The developed DyCeModel tool is quite flexible; it provides embedding, addition, and mixing of existing mathematical models. If you have experimental data of interest substance, it is possible to automatically select unknown parameters, which speeds up work with the model and makes it more stable. As an output, DyCeModel produces a statistical summary of cellular composition that can be used to predict the effect of hormones on cell proliferation activity. The resulting models provide a promising framework for further study of the role of hormone-controlled gene regulatory networks in cell dynamics.

Funding: The work was supported by the budget project FWNR-2022-0020.

References

1. Nikolaev S.V., Kolchanov N.A., Fadeev S.I., Kogai V.V., Mjolsness E.A. Study of a one-dimensional model for control of the renewing zone size in biological tissue. *Vychislitel'nyye Tekhnologii*. 2006;11(2):67-81 (in Russian)
2. Band L.R., Wells D.M., Fozard J.A., Ghetiu T. et al. Systems analysis of auxin transport in the arabidopsis root apex. *Plant Cell*. 2014;26(3):862-875. doi 10.1105/tpc.113.119495
3. De Rybel B., Adibi M., Breda A.S. et al. Integration of growth and patterning during vascular tissue formation in *Arabidopsis*. *Science*. 2014;345(6197):1255215. doi 10.1126/science.1255215
4. Dubreuil C., Jin X., Grönlund A., Fischer U. A Local Auxin Gradient Regulates Root Cap Self-Renewal and Size Homeostasis. *Curr Biol*. 2018;28(16):2581-2587.e3. doi 10.1016/j.cub.2018.05.090
5. Marconi M., Gallemi M., Benkova E., Wabnik K. A coupled mechano-biochemical model for cell polarity guided anisotropic root growth. *Elife*. 2021;10: e72132. doi 10.7554/eLife.72132
6. Likhoshvai V., Ratushny A. Generalized Hill function method for modeling molecular processes. *J Bioinform Comput Biol*. 2007;5(02b):521-531

doi 10.18699/bgrs2024-06-31

In silico characterization of Guar genes homologous to soybean *E1–E4* maturity genes

Criollo L.*, Zamalutdinov A., Potokina E.
The Project Center for Agro Technologies, Skolkovo Institute of Science and Technology, Moscow, Russia
** Luisa.Criollo@skoltech.ru*

Key words: guar; photoperiod sensitivity; flowering time; homologous soybean genes

Motivation and Aim: Guar (*Cyamopsis tetragonoloba* (L.) Taub.) is an economically important short-day legume, whose seeds are the main source of Guar gum, a compound used as a thickening agent in the food industry and as a fracturing fluid during oil and gas extraction [1]. In Russia the main obstacle to its cultivation is the long photoperiod during summer which delays flowering in guar [2]. In other legumes such as Soybean, the flowering time is mostly influenced by four maturity genes, denominated as E1, E2, E3 and E4 [3]. However, in Guar these genes have not been characterized, which creates an issue for the implementation of genome editing technologies in these flowering-related genes. In this study, we used a bioinformatics approach to characterize the Guar genes homologous to soybean E1–E4 maturity genes.
Methods and Algorithms: The nucleotide sequences of the soybean E1–E4 maturity genes were retrieved from NCBI. Subsequently, BLAST was used to perform a sequence similarity search in the guar genome and identify homologous sequences. The genomic sequences with high percentage of homology were retrieved and gene prediction tools such as AUGUSTUS or FGENESH were used to predict the exon-intron structure of the candidate genes. MEGA 11 was used to carry out sequence alignment and identify conserved regions. Functional annotation was performed using InterProScan and I-TASSER server. The candidate genes were amplified through PCR in Guar accessions with different flowering time and sequenced using Sanger method. Additionally, transcriptome data was used to confirm the expression of the E1–E4 homologous genes in Guar.
Results: BLAST search against the Guar genome allowed the identification of homologous sequences to the soybean E1–E4 genes. Sequence alignment between Guar homologs and soybean genes revealed high sequence similarity and several conserved regions, indicative of the presence of conserved structural features. The gene structure prediction showed similar exon-intron structure between Guar candidate genes and their respective soybean counterparts. Functional annotation showed putative protein domains in the Guar homologs, providing a better understanding of their potential biological function. Sequence analysis of the Guar homologs in early and late flowering accessions provided a first insight into the presence of potential SNP associated with flowering time regulation. Finally, analysis of the transcriptome data confirmed the expression of the E1–E4 homologous genes in Guar.
Conclusion: The Guar genes homologous to soybean E1–E4 maturity genes were identified and characterized using several bioinformatic tools. Analysis of the sequence of Guar homologs in early and late flowering accessions provided a first insight into the presence of potential SNP associated with flowering time regulation. The *in silico* characterization of these Guar genes lays the foundation for further functional studies

and genome editing experiments targeting these genes to analyze the genetic network that underlies the flowering time regulation in Guar.

Funding: The study is supported by Russian Science Foundation grant No. 24-26-00073.

References

1. Hasan A.M., Abdel-Raouf M.E. Applications of guar gum and its derivatives in petroleum industry: A review. *Egypt J Pet*. 2018;27(4):1043-1050. doi 10.1016/j.ejpe.2018.03.005
2. Grigoreva E., Tkachenko A., Arkhimandritova S. et al. Identification of Key Metabolic Pathways and Biomarkers Underlying Flowering Time of Guar (*Cyamopsis tetragonoloba* (L.) Taub.) via Integrated Transcriptome-Metabolome Analysis. *Genes*. 2021;12(7):952. doi 10.3390/genes12070952
3. Liu L., Song W., Wang L. et al. Allele combinations of maturity genes E1-E4 affect adaptation of soybean to diverse geographic regions and farming systems in China. *PLoS One*. 2020;15(7):e0235397. doi 10.1371/journal.pone.0235397

doi 10.18699/bgrs2024-06-32

GWAS using the IIIVmrMLM model provides new genomic associations with agronomic traits in Chickpea

Duk M.[1, 2]*, Gurkina M.V.[3], Nekrasov A.Y.[4]

[1] *Peter the Great St.Petersburg Polytechnic University, St.Perersburg, Russia*

[2] *Ioffe Institute, RAS, St. Petersburg, Russia*

[3] *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, Astrakhan Experiment Station of VIR, Astrakhan, Russia*

[4] *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, Kuban Experiment Station of VIR, Krasnodar, Russia*

* *duk@mail.ioffe.ru*

Key words: chickpea; *Cicer arietinum*; GWAS; IIIVmrMLM; favorable alleles

Motivation and Aim: Chickpea is one of the most important crops and is cultivated both for food (kabuli type) and animal feed (desi type) [1]. Although it is grown in many countries and used in many traditional Asian dishes, the genetic diversity of this crop has been relatively reduced due to domestication. Due to this, chickpea plants are susceptible to various abiotic stresses [2] and diseases. In this regard, the search for new variants associated with different agronomic traits is important for breeding programs. In addition, agronomic traits in chickpea vary greatly under different environmental conditions, so traditional genome-wide association searching methods yield different variants in different environments that may not be meaningful in other environments. The new IIIVmrMLM genomic variant search method [3] allows it to be used in multiple environments as well as in each environment separately. In addition, GWAS methods based on the MLM algorithm have a Bonferroni correction for multiple testing that is too stringent to find associations with complex traits in crop [4]. The IIIVmrMLM method solves this problem by estimating effects using a single multilocus model and then by estimating non-zero effects using the likelihood ratio test.

Methods and Algorithms: The collection of the All-Russian Institute of Plant Genetic Resources (VIR), which we used in our study, consisted of 171 chickpea accessions, of which 147 accessions were landraces. We considered 12 phenotypic traits measured in two regions of Russia, Astrakhan (46°06′N, 48°04′E, altitude 24 m) and Kuban (45°18′N, 40°52′E, altitude 138.9 m) in 2022. We used the IIIVmrMLM package [5] in Multi_env mode to search in two environments and in Single_env mode to search for variants in each environment separately. Phenotypic data were quantile normalized for each trait. We considered phenotypic traits such as: days from emergence to start of flowering, flowering duration in days, days from emergence to full maturity, plant height, height to first pod, weight of all pods per plant, number of all pods per plant, weight of plant with pods (biological yield), weight of one hundred seeds, number of primary and secondary branches and leaf size.

Results: We found a total of 233 QTNs and 204 QEIs (QTN-environment interactions) for 12 traits using IIIVmrMLM in Multi_env mode and 455 associations in total for each of locations using a Single_env mode. Associations were located in all 8 chromosomes. QTNs found for each of environments explained 36.56 % to 55.5 % of the variation in Astrakhan and 36.32 % to 53.26 % of the variation in Kuban. In the Multi_env mode,

QTN explained about 27.5 % of the variation, whereas QEI explained on average 38.8 % of the variation, ranging from 24 % to 52.4 % for the different traits. The 29 QTNs found using Single_env model at each location were previously found by the FarmCPU, Blink, SUPER, MLMM, and MLM methods using the GAPIT package [6, 7], whereas only 8 QTNs and 5 QEIs using Multi_env model were found previously by other methods. The 43 QTNs and 40 QEIs found with Multi_env model, as well as 63 QTNs predicted with Single_env model, were located within or around genes. There were only 9 overlaps between the lists of these genes.

In addition, we found favorable alleles for such traits as days from emergence to maturity, weight of the plant with pods, weight and number of pods per plant. The IIIVmrMLM method allowed us to estimate the additive effect of each variant in each location, which made it possible to find accessions with a good combination of favorable alleles for Astrakhan and Kuban. As a result, we found 5 landraces with good combinations of QTN and QEI for Kuban and two landraces with good combinations of QTN and QEI for Astrakhan.

Conclusion: When IIIVmrMLM was applied to data from single or multiple environments, the results obtained differed significantly. The Single_env model applied separately to data from each location had more overlap with GWAS results found with GAPIT. However, such models did not take into account G×E interaction and may result in associations which cannot be reproduced in another climatic conditions. The Multi-env model implemented in IIIvmrMLM estimates effects of genes and all types of interactions between alleles, genes and environments. The environment-specific alleles of the QEI loci are of particular interest, as they can be used by breeders to obtain varieties with improved adaptation to specific climatic conditions.

Funding: The study is supported No. 22-46-02004.

References

1. Varshney R.K. et.al. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat Genet*. 2019;51(5):857-864. doi 10.1038/s41588-019-0401-3
2. Jha U.C., Kole P.C., Singh N.P. QTL mapping for heat stress tolerance in chickpea (*Cicer arietinum* L.). *Legume Res.* 2019;44(4):382-387. doi 10.18805/LR-4121
3. Zhang Y.-W., Tamba C.L., Wen Y.-J. et al. MrMLM v4.0.2: An R Platform for Multi-Locus Genome-Wide Association Studies. *Genomics Proteomics Bioinformatics*. 2020;18(4):481-487. doi 10.1016/j.gpb.2020.06.006
4. Zhang Y.-M., Jia Z., Dunwell J.M. Editorial: The Applications of New Multi-Locus GWAS Methodologies in the Genetic Dissection of Complex Traits. *Front Plant Sci*. 2019;10:100. doi 10.3389/fpls.2019.00100
5. Li M., Zhang Y.-W., Xiang Y., Liu M.-H., Zhang Y.-M. IIIVmrMLM: The R and C++ Tools Associated with 3VmrMLM, a Comprehensive GWAS Method for Dissecting Quantitative Traits. *Mol Plant*. 2022;15:1251-1253. doi 10.1016/j.molp.2022.06.002
6. Wang J., Zhang Z. GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinformatics.* 2021;19(4):629-640. doi 10.1016/j.gpb.2021.08.005
7. Duk M.A., Kanapin A.A., Bankin M.P., Vishnyakova M.A., Bulyntsev S.V., Samsonova M.G. Genome-Wide Association Analysis in Chickpea Landraces and Cultivars. *Biophysics*. 2023;68(6):952-963

doi 10.18699/bgrs2024-06-33

Application of mathematical modeling to analyze the flowering activation motif during vernalization in legumes

Duk M.[1, 2], Gursky V.[2], Samsonova M.[1], Surkova S.[1]*

[1] *Mathematical Biology and Bioinformatics Laboratory, Peter the Great Saint Petersburg Polytechnic University, St. Petersburg, Russia*

[2] *Theoretical Department, Ioffe Institute, RAS, St. Petersburg, Russia*

* *surkova_syu@spbstu.ru*

Key words: legumes; flowering; vernalization; gene networks; dynamical model; feed-forward loop

Motivation and Aim: The mechanism of vernalization-induced flowering has been extensively studied in *Arabidopsis* but remains largely unknown in legumes. The orthologs of the *FLOWERING LOCUS C (FLC)* gene, a major regulator of vernalization response in *Arabidopsis*, are absent or non-functional in the vernalization-sensitive legume species. Nevertheless, the legume integrator genes *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) have been shown to be involved in the vernalization-induced floral transition. However, their regulatory contribution remains unexplored. The *FT* and *SOC1* genes are usually present in several paralogous copies in the legume genomes. We performed the theoretical and data-driven analyses of a feed-forward regulatory motif that includes a vernalization-responsive *FT* gene and several *SOC1* genes, which independently activate the meristem identity gene *PROLIFERATING INFLORESCENCE MERISTEM* (*PIM*) and thereby mediate floral transition [1].

Methods and Algorithms: The model is formulated in terms of the ordinary differential equations. The dynamical models to fit the experimental data are based on the kinetic equations with the Michaelis–Menten kinetics.

Results: Our theoretical model showed that the multiple regulatory branches in the regulatory motif facilitated the elimination of meaningless signals and amplified useful signals from the upstream regulator. We further developed and analyzed four data-driven models of *PIM* activation in *Medicago trancatula* in vernalized and non-vernalized conditions in wild type and *fta1-1* mutants [2]. The model with *FTa1* providing both direct activation and indirect activation via three intermediate activators, *SOC1a*, *SOC1b*, and *SOC1c*, resulted in the most relevant *PIM* dynamics. In this model, the difference between regulatory inputs of *SOC1* genes was nonessential. As a result, in the *M. trancatula* model, the cumulative action of *SOC1a*, *SOC1b*, and *SOC1c* was favored [1]. This is in line with the suggested hypothesis of functional redundancy of *SOC1* genes.

Conclusion: Overall, we first performed the in silico analysis of vernalization-induced flowering in legumes. The considered vernalization network motif can be supplemented with additional regulatory branches as new experimental data become available.

Funding: The study is supported by Russian Science Foundation (RSF), grant No. 23-26-00203.

References

1. Duk M.A., Gursky V.V., Samsonova M.G., Surkova S.Y. Modeling the flowering activation motif during vernalization in legumes: a case study of *M. trancatula*. *Life.* 2024;14:26. doi 10.3390/life14010026
2. Fudge J.B., Lee R.H., Laurie R.E., Mysore K.S., Wen J., Weller J.L., Macknight R.C. *Medicago truncatula SOC1* genes are up-regulated by environmental cues that promote flowering. *Front Plant Sci*. 2018;27(9):496. doi 10.3389/fpls.2018.00496

doi 10.18699/bgrs2024-06-34

Applications of artificial intelligence, machine learning, and deep learning in plant breeding

Eftekhari M.[1]*, Ma C.[2, 3]**, Orlov Y.L.[4, 5]***

[1] *Department of Horticultural Sciences, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran*

[2] *State Key Laboratory of Crop Stress Resistance and High-Efficiency Production, Center of Bioinformatics, College of Life Sciences, Northwest A&F University, Shaanxi, China*

[3] *Key Laboratory of Biology and Genetics Improvement of Maize in Arid Area of Northwest Region, Ministry of Agriculture, Northwest A&F University, Shaanxi, China*

[4] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[5] *Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia*

** m.eftekhari@modares.ac.ir; ** chuangma2006@gmail.com; *** orlov@bionet.nsc.ru*

Key words: computational plant biology; AI-driven plant breeding; omics data analysis; crop resilience; disease detection; food security

Motivation and Aim: In recent years, the field of plant breeding has witnessed a paradigm shift driven by advancements in artificial intelligence (AI) technologies, including machine learning (ML) and deep learning (DL) technologies. These cutting-edge techniques have transformed our understanding of plant biology. From decoding the intricate molecular mechanisms of plant defense to automating disease detection and optimizing nutrient levels, AI is reshaping the landscape of plant breeding. AI-assisted omics techniques offer insights into plant-pathogen interactions and facilitate the identification of stress-responsive genes [1, 2].

Methods and Algorithms: We have organized thematic journal issue at Frontiers in Plant Science – Research Topic "Applications of artificial intelligence, machine learning, and deep learning in plant breeding" [3]. We overview the papers published in this Research Topic on the application of computer techniques in plant science.

Results: We collected research papers on the topic of AI applications in plant biology in areas of sequencing data analysis, image recognition, technology process optimization [3]. Murmu et al. (https://doi.org/10.3389/fpls.2024.1292054) highlighted the potential of AI algorithms, particularly ML and DL, in decoding complex omics data to elucidate the molecular foundations of plant defense. The authors explored AI-assisted omics techniques' applications, challenges, and prospects in enhancing crop protection strategies and ensuring global food security amidst environmental challenges. By integrating AI with omics technologies, researchers can unravel intricate gene regulatory networks and develop targeted interventions for enhancing crop resilience. As we confront the challenges of climate change and emerging diseases, AI-driven approaches offer a robust toolkit for ensuring global food security and sustainability in agriculture. Climate change poses significant threats to agricultural systems, emphasizing the importance of elucidating cold defense mechanisms in crops. Other authors in the Research Topic introduced the Self Organizing Maps (SOM)-based ML method to decipher gene expression patterns in response to different temperature regimes [3]. The YOLO (You Only Look Once) architecture, known for its real-time object detection capabilities, is employed for object detection in plant image analysis [4]. Li et al. (https://doi.org/10.3389/fpls.2023.1289692) showed the efficacy of the CFNet-

VoVGCSP-LSKNet-YOLOv8s model in accurately identifying cotton pests and diseases amidst challenging environmental conditions. The model's superior performance offers a promising solution for real-time monitoring and early intervention in pest and disease outbreaks. This article heralds a new era in cotton plant breeding, wherein cutting-edge AI, ML, and DL techniques converge to address age-old challenges with remarkable precision and efficiency. A novel DL-based architecture, DeepPlantNet, for efficient and accurate prediction and categorization of plant leaf diseases was introduced. Another key phenotypic trait in plants – pubescence, correlates with stress resistance, particularly in wheat. Artemenko et al. [5] proposed an AI-driven approach using CNNs to automate glume pubescence detection, addressing the limitations of traditional methods and enhancing breeding efficiency. Their method offers a reliable and efficient solution for phenotype analysis, empowering breeders with advanced tools for cultivar selection and stress resilience enhancement.

Conclusion: The analysis of molecular mechanisms underlying plant adaptation to environmental changes and stress response is crucial for plant biotechnology. The key approaches include bioinformatics methods, high-throughput sequencing, and post-genome technologies. Genomics and bioinformatics facilitate the modeling of protein–protein and gene regulatory interactions in plant cells, providing a basis for better crop production and sustainability. Overall, AI, ML, and DL techniques offer unique opportunities from deciphering complex omics data to automating phenotypic trait analysis and disease detection to revolutionize breeding practices, develop stress-tolerant and high-yielding crop varieties, and contribute to global food security in the face of escalating environmental challenges [2]. Continued investment in AI applications in plant breeding holds the key to unlocking the full potential of agriculture. We aim to continue a series of Research Topics (special journal issues) on computational plant biology and bioinformatics application [6] at different publishers' platforms.

Funding: The study is supported by Russian Science Foundation (grant project "Smart Crop – Cognitive Platform for Reconstruction, Visualization and Analysis of Stress Response Networks based on ANDSystem and Multiomics in Rice and Wheat", No. 23-44-00030).

References

1. Mahmood U., Li X., Fan Y. et al. Multi-omics revolution to promote plant breeding efficiency. *Front Plant Sci.* 2022;13:1062952. doi 10.3389/fpls.2022.1062952
2. Chao H., Zhang S., Hu Y. et al. Integrating omics databases for enhanced crop breeding. *J Integr Bioinform*. 2023;20(4):20230012. doi 10.1515/jib-2023-0012
3. Eftekhari M., Ma C., Orlov Y.L. Editorial: Applications of artificial intelligence, machine learning, and deep learning in plant breeding. *Front Plant Sci.* 2024;15:1420938. doi 10.3389/fpls.2024.1420938
4. Liu Y., Zhao Q., Wang X., Sheng Y., Tian W., Ren Y. A tree species classification model based on improved YOLOv7 for shelterbelts. *Front Plant Sci.* 2024;14:1265025. doi 10.3389/fpls.2023.1265025
5. Artemenko N.V., Genaev M.A., Epifanov R.U. et al. Image-based classification of wheat spikes by glume pubescence using convolutional neural networks. *Front Plant Sci.* 2024;14:1336192. doi 10.3389/fpls.2023.1336192
6. Orlov Y.L., Chen M. Special issue on "Plant biology and biotechnology: focus on genomics and bioinformatics 2.0". *Int J Mol Sci.* 2023;24:17588. doi 10.3390/ijms242417588

doi 10.18699/bgrs2024-06-35

Development of genetic models for validation of candidate genes associated with cold stress response in tea plant

Egorova A.[1, 2]*, Fizikova A.[1, 3], Fomin I.[2], Kostina N.[2], Gerasimova S.[1, 2], Samarina L.[1, 3], Malukova L.[1]

[1] *Federal Research Centre the Subtropical Scientific Centre, RAS, Sochi, Russia*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[3] *Sirius University of Science and Technology, Sirius, Russia*
* *egorova@bionet.nsc.ru*

Key words: *Camellia sinensis*; Cas9/gRNA; protoplasts; genetic transformation

Motivation and Aim: Tea (*Camellia sinensis* (L.) Kuntze) is a perennial woody plant, one of the most important agricultural crops for human consumption, medicinal and functional purposes, grown in more than 60 countries worldwide. Cold is the main factor limiting the expansion of tea plantations to more northern latitudes and limiting their productivity. However, knowledge of cold response in woody plants, including tea, is fragmented and the genetic basis of cold tolerance requires further investigation. The genetic diversity of tea in the Northwest Caucasus can serve as a valuable source of cold tolerant germplasm for the development of new tolerant cultivars. Kolkhida is the best local large leaf tea cultivar, known for its high-quality leaves used in black and green tea production. Transcriptional analysis of cv. Kolkhida identified candidate genes involved in long-term cold stress responses [1]. The aim of this study is to functionally validate these candidate genes using knockout and overexpression methods.

Methods and Algorithms: Cas9/gRNA-mediated site-directed mutagenesis was used to knock out candidate genes in cv. Kolkhida. Overexpression of candidate genes was performed in *Nicotiana tabacum* SR1 plants.

Results: Two candidate genes (*COR413PM1-like* and *ELIP1*) were selected for functional validation. Two target sites were selected in the coding sequences of each gene. Cas9/gRNA vectors were constructed for the target sites in these genes. The extraction and transformation of protoplasts from young tea leaves was established and the mutagenic activity of the vectors was demonstrated in protoplasts. The cas9/gRNA-carrying cassettes were then introduced into the binary vectors which are planned to be used for stable transformation of tea and subsequent regeneration of plants carrying mutations in the two target genes.

The candidate genes were cloned from the cv. Kolkhida and overexpression vectors were constructed. *Agrobacterium*-mediated genetic transformation of *N. tabacum* was performed, resulting in ten transgenic lines for each gene. Based on qPCR analysis, five lines with the highest expression for each gene were selected and T1 generations were obtained.

Conclusion: During this work, the development of genetic models was initiated, which will be used to study the functions of candidate genes in response to cold stress.

Funding: The study is supported by a grant from the Russian Science Foundation (No. 23-46-00002).

References

1. Samarina L., Wang S., Malyukova L. et al. Long-term cold, freezing and drought: Overlapping and specific regulatory mechanisms and signal transduction in tea plant (*Camellia sinensis* (L.) Kuntze). *Front Plant Sci.* 2023;14:1145793. doi 10.3389/fpls.2023.1145793

doi 10.18699/bgrs2024-06-36

A unified data preprocessing framework to address inconsistencies in comparative plastid genome studies

Emirsaliev A.[1, 2, 3]*, Salina E.[1, 2] , Mitrofanova I.[3], Afonnikov D. [1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia;*

[3] *N.V. Tsitsin Main Botanical Garden, RAS, Moscow, Russia*

* *emirsaleh@bionet.nsc.ru*

Key words: plastome; chloroplast genome; data standardization; data harmonization; normalization; comparative genomics; phylogenetics; phylogenomics

Motivation and Aim: Plastid genomes (plastome) are the most underused data sources for phylogenetic analysis of plants. There are also a lot of widely used universal barcodes for plant genotyping which are derived from plastid genome data. With the rise of high throughput sequencing technologies ever increasing number of the complete plastid genome sequences greatly expanded the amount of available data to be analyzed. After the first complete plastome sequence was published [1] thousands of complete organellar genomes of plants were obtained by a global research community. In the long historical way of new plastid genomes assembling, describing and analyzing, diverse approaches, tools and standards were made use of at different time periods. As a result, the corpus of data obtained may be inconsistent due to different methods used in data processing. Such a differences affect the feature names, sequences and annotations of the records deposited in databases. The analysis tools used in comparative genome studies expect the data to be comparable, which means for the sequences to be of proper linear representation, and for features to be annotated and named according to standards. The sequence splits, as well as the orientation and order of large substrings also referenced as genome regions might affect the alignment scores. Inconsistent naming greatly impacts the records to be retrieved from databases and annotations to be involved in analysis. Some analyses fail when features are not properly named and ordered. We argue for benefits of *a priori* plastome data standardization to eliminate shortcomings related to the origin of data being analyzed. The aim of this work is to present a unified data preprocessing framework that standardizes plastid genome data, enabling more accurate and meaningful comparative analyses, and advancing our understanding of plant genomics and phylogeny.

Methods and Algorithms: The proposed framework handles common sources of discrepancies in plastid genome data through the following stages:

1. Data querying and retrieval from Nucleotide Database of NCBI/EMBL-EBI/DDBJ via Entrez API.
2. Records deduplication.
3. Assessing the coverage of annotation.
4. Validating the linearized representation of plastome sequences.
5. Defining the records to be kept.
6. Properly ordering and orienting genome regions.
7. Filling missing but required fields in the annotation.
8. Extracting data for analysis based on the purpose.

The framework is implemented in an interactive Jupyter Notebook environment, orchestrating Biopython, Pandas, and external tools like cpDNA_standardization [2], GeSeq [3], and NOVOWrap [4] for validation and annotation. One can easily configure each step if needed by modifying particular cell in the notebook.

Results: Here we describe a framework addressing some data quality issues, enabling more accurate and meaningful comparative analyses of plastid genomes, potentially yielding novel insights and advancing our understanding of plant genomics and phylogeny. The steps formalized in the framework allows to prepare plastome data for comparative genome studies, addressing some common issues and providing control at stages that might affect the analysis quality, like deduplication, fixing naming issues and sequences regions shifting, reordering and reorienting. The framework handles issues occuring systematically in the plastome data that were empirically identified while analyzing sample plastome data of the Cichorieae tribe (Asteraceae family, Eudicots). The framework documentation and code, as well as sample data, are available at github.com/asan-emirsaleh/plastome-preprocessing.

Conclusion: The proposed unified data preprocessing framework aims to eliminate shortcomings related to the use of diverse tools for data preparation and annotation before depositing in databases, by standardizing and ensuring data comparability. With addressing inconsistencies in data processing, the framework facilitates reliable and reproducible comparative genomic studies. Furthermore, this approach can potentially be extended to other small genomes of complex structure which fit the circular model, like mitochondrion genomes. The framework could also be integrated in plastome data analysis pipelines.

Funding: The study is supported by the budget project No. FWNR-2022-0017.

References

1. Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N., Matsubayashi T., Zaita N., Chunwongse J., Obokata J., Yamaguchi-Shinozaki K., Ohto C., Torazawa K., Meng B.Y., Sugita M., Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A., Tohdoh N., Shimada H., Sugiura M. The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J*. 1986;5(9):2043-2049. doi 10.1002/j.1460-2075.1986.tb04464.x
2. Turudić A., Liber Z., Grdiša M., Jakše J., Varga F., Šatović Z. Towards the Well-Tempered Chloroplast DNA Sequences. *Plants*. 2021;10(7):1360. doi 10.3390/plants10071360
3. Tillich M., Lehwark P., Pellizzer T., Ulbricht-Jones E.S., Fischer A., Bock R., Greiner S. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6-W11. doi 10.1093/nar/gkx391
4. Wu P., Xu C., Chen H., Yang J., Zhang X., Zhou S. NOVOWrap: An automated solution for plastid genome assembly and structure standardization. *Mol. Ecol. Resour*. 2021;21(6):2177-2186. doi 10.1111/1755-0998.13410

doi 10.18699/bgrs2024-06-37

Generation of semi-dwarf tropical maize by multiplex genome editing

Gerasimova S.V.[1, 2]*, Pinto M.S.[1], Nonato J.[1], Altoe I.[1], Yassitepe J.E.C.T.[1, 3], Dante R.A.[1, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Genomics for Climate Change Research Center (GCCRC), Universidade Estadual de Campinas, Campinas, Brazil*

[3] *Embrapa Agricultura Digital, Campinas, Brazil*

* *sophia.v.gerasimova@gmail.com*

Key words: CRISPR/Cas; gibberrelin; GA20ox; plant biotechnology; breeding

Motivation and Aim: Maize is the most important modern agricultural crop. A significant portion of global maize production comes from tropical regions. Tropical maize is better adapted to heat and drought than varieties from temperate regions, but it is characterized by lower productivity. The height of plants in many tropical genotypes is significantly greater than that of elite material from temperate climates, which is associated with lower yield and less resistance to lodging. Reducing the height of maize plants is one of the most pressing tasks in breeding this crop. Until recently, all methods of improving tropical corn varieties were limited to classical breeding. However, we have recently developed a genome editing method for tropical maize [1], which allows the application of targeted gene modification methods to enhance the most important traits. Gibberellin (GA) 20-oxidase (GA20ox) genes are known to control plant height. In this study, genome editing approach was applied to the GA20ox gene family in few tropical maize lines with the aim of reducing plant height.

Methods and Algorithms: Few guide RNAs (gRNAs) were designed for directing Cas9 simultaneously to multiple genes of the maize Ga20ox gene family. Cas9/gRNA expressing cassette together with *Wus/Bbm* morphogenic genes was coupled with ternary vector system for increasing the efficiency of Agrobacterium-mediated transformation. T0 and T1 plant generations were genotyped by Sanger sequencing. Combination of digital and manual phenotyping was performed to assess morphological characteristics of individual plants in T1. Homozygous mutant lines were established by self-pollination of selected individual mutants.

Results: Modification of multiple genes of the Ga20ox gene family was performed on three tropical maize lines from the CIMMYT collection and temperate model line B104. In all lines, a wide range of phenotypes in the T0 generation associated with mutations in the target genes was obtained. Some T0 regenerant plants from two tropical lines were used to produce the next generation through self-pollination. In the T1 generation, phenotypes obtained in the T0 generation were reproduced, and several new phenotypes associated with changes in plant height and shape were also identified.

Conclusion: The use of the most advanced genome editing methods enables overcoming genotype-specific transformation barriers and obtaining modifications in exotic genotypes of crop plants. Modification of genes in the gibberellic acid regulatory pathway allows obtaining a spectrum of morphological changes potentially valuable for breeding.

Funding: This work was funded by grant 2016/23218-0 “Genomics for Climate Change Research Center (GCCRC)” from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). SG is supported by a visiting researcher grant from FAPESP (2022/09418-7).

References

1. Hernandes-Lopes J., Pinto M.S., Vieira L.R. et al. Enabling genome editing in tropical maize lines through an improved, morphogenic regulator-assisted transformation protocol. *Front Genome Ed.* 2024;5:2673-3439. doi 10.3389/fgeed.2023.1241035

doi 10.18699/bgrs2024-06-38

Identification of subpopulation in Russian soybean germplasm collection using advanced genetic methods

Habib Y.[1]*, Zamalutdinov A.[1], Boldyrev S.[1], Schegolkov A.[2], Ben C.[1], Gentzbittel L.[1]

[1] *Skolkovo Institute of Science and Technology, Moscow, Russia*

[2] *SOKO, Krasnodar, Russia*

* *Yawar.Habib@skoltech.ru*

Key words: soybean germplasm; admixture; Discriminant Analysis of Principal Components (DAPC)

Motivation and Aim: Globally recognized as a crucial legume, soybean (*Glycine max* (L.) Merr.) is esteemed for its high protein and oil content, serving as a key component in diverse food products. The foundation of successful breeding programs requires a comprehensive understanding of the genetic variation and population dynamics existing within contemporary germplasm repositories. Assessing genetic diversity and identifying the number of subpopulations is crucial for enhancing breeding strategies and preserving genetic resources. This study aims to employ sophisticated techniques such as Admixture, Discriminant Analysis of Principal Components (DAPC), integrated with K-means clustering, to identify various subgroups within Russian soybean collections and elucidate their genetic diversity and population structure.

Methods and Algorithms: SNP calling for GBS data was performed using "ref_map.pl" from the Stacks software package with default parameters, while for WGS data, GATK 4.2.6.1 Haplotype Caller was used [1]. Quality control was conducted with PLINK v1.07 [2], using thresholds for missing genotype and individual rates at 0.6 and 0.01 respectively, MAF at 0.05. To remove SNPs in linkage disequilibrium, haplotype blocks were constructed using the D' method [3] with PLINK v1.07, defining strong LD by a 95 % confidence bound on D' > 0.98 and < 0.9 for recombination evidence, with a 200kb window size. Admixture patterns were analyzed using the software ADMIXTURE v1.3 [4] with 50 bootstrap replicates, and cross-validation was done to determine the optimal K value. DAPC, implemented in R's adegenet 2.0 package, was used to identify genetic clusters via K-means clustering and Bayesian Information Criterion (BIC), transforming data with PCA and performing Discriminant Analysis on retained PCs using cross-validation.

Results: In this study, a set of 59,830 SNPs, genotyped from 344 soybean accessions (50 by WGS and 294 by GBS), was utilized. Following quality control procedures, 25,388 SNPs were retained across 323 soybean genotypes, achieving a genotyping rate of 93.58 % after excluding SNPs and individuals that did not meet the criteria for missingness and minor allele frequency threshold. The number of haplotype blocks computed were 3,225 and two SNPs from each block were selected based on higher MAF and the maximum distance among each other. The lowest value of cross validation error was associated with K = 15 showing in Fig. 1(a). Fig. 1(b) displays the proportion of ancestral heritage for these 15 groups as determined by admixture analysis. Many individuals display homogenous ancestry, with bars of a single color indicating over 90 % genetic material from one ancestral group. Clusters of individuals from 250 to 268 and 288 to 307 exemplify this, having similar parental lines. In total, 93 individuals showed homogenous ancestry. Conversely, some individuals have highly mixed

ancestry. For instance, individual 208's highest single ancestry proportion is only 17.5 %, while individual 203's is 18.7 %. Others, like individuals 221, 205, 135, and 202, have single ancestry proportions ranging from 18.9 % to 21.5 %, reflecting diverse genetic contributions.

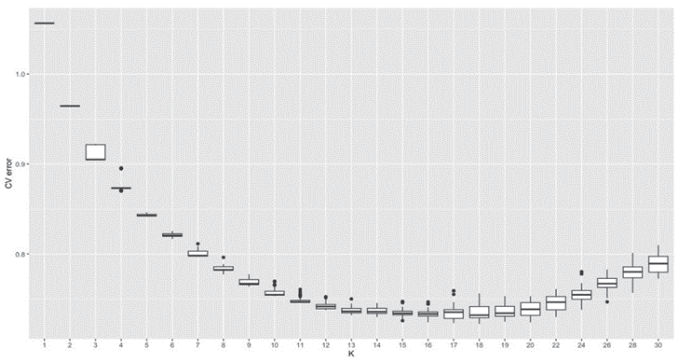


Fig. 1. (a) Cross-validation error for each value of K, indicating the optimal number of ancestral populations

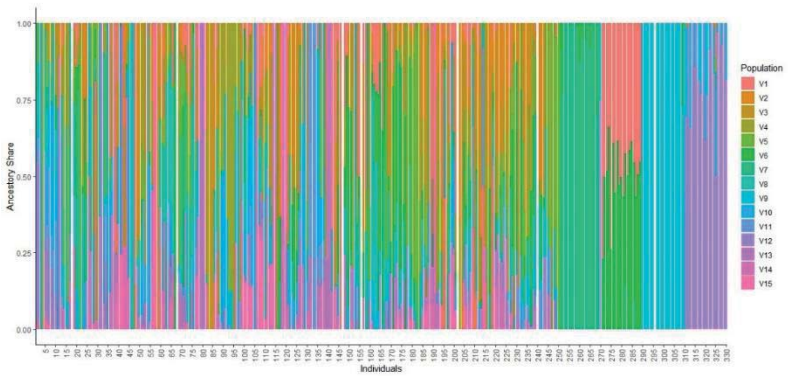


Fig. 1. (b) Ancestry proportions for each individual in the study, as inferred from the Admixture analysis with K=15 (15 ancestral populations)

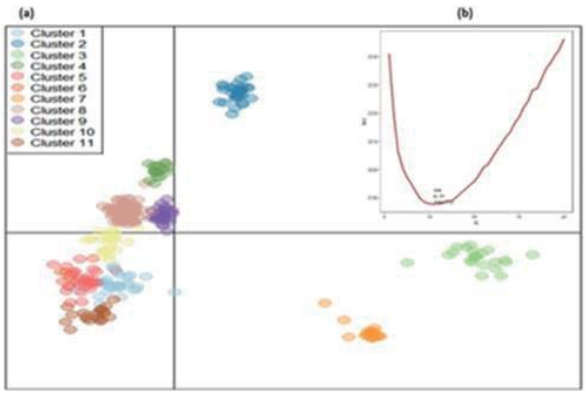


Fig. 2. (a) Shows Discriminant Analysis of Principal components (DAPC) scatter plots with PC1 and PC2 for identification of 11 unique soybean groups with embedded fig. 2 (b) which shows K means clustering visualization based on Bayesian Information Criteria (BIC)

The valley value plot in Fig 2(b) for K means clustering shows an elbow at K = 11, which will be considered the best number of clusters for this soybean germplasm collection. The scatter plot in Fig. 1(a) from the DAPC shows these 11 distinct groups, with 30 principal components and 6 discriminant functions used to describe how these clusters are related. Clusters 2 and 7 are densely packed with genetically similar lines from specific parent crosses, as seen in the admixture analysis. This dense clustering confirms the quality of the genetic data and the accuracy of the clustering. In contrast, the less dense clusters show more genetic diversity, which is good for the genetic diversity of the germplasm. These diverse clusters are important because they might possess different traits valuable for different breeding targets. The number of accessions in each subpopulation varied significantly. Subpopulation VIII had the highest number with 70 accessions, while Subpopulation III and XI had the lowest numbers, with 5 and 20 accessions, respectively. The distribution aids in elucidating the genetic architecture of the soybean germplasm, serving as a guide for breeding and conservation initiatives.
Conclusion: This study mapped out the genetic makeup of Russian soybean germplasm, revealing 11 subpopulations using DAPC and 15 ancestral populations through admixture analysis. Despite the differences in the number of clusters, there was notable consistency in individual clustering, particularly in densely populated clusters like Clusters 2 and 7, which showed high genetic similarity. The genetic diversity observed among subpopulations is crucial for the preservation of germplasm and for guiding breeding programs focused on developing new cultivars. These findings underscore the importance of integrating different analytical approaches to obtain comprehensive insights into genetic diversity and population structure, with important implications for future studies on association mapping and crop improvement.

References

1. O'Connor B.D., Van der Auwera G. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media, 2020
2. Purcell S., Neale B., Todd-Brown K. et al. PLINK: A Toolset for Whole Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007;81(3):559-575
3. Gabriel S.B., Schaffner S.F., Nguyen H. et al. The structure of haplotype blocks in the human genome. *Science*. 2002;296(5576):2225-2229
4. Alexander D.H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655-1664

doi 10.18699/bgrs2024-06-39

Assembly and analysis of plastomes for 30 potato cultivars grown in Russia

Karetnikov D.I.[1, 2]*, Salina E.A.[1, 2], Kochetov A.V.[1, 2], Afonnikov D.A.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *karetnikovmit@bionet.nsc.ru*

Key words: potato; *Solanum tuberosum* L.; chloroplast DNA; cytoplasmic DNA type; genetic diversity; SNP; InDels

Motivation and Aim: The chloroplast DNA of *Solanum tuberosum* L. is characterized by the isolation of different types of cytoplasmic DNA. There are five main types of potato chloroplast DNA: W, T, A, C, S. Each type is characterized by its mutation in the plastome. Thus, the T-type chloroplast DNA resulted from the deletion of 241 nucleotides in the primitive W-type chloroplast DNA. Determination of potato chloroplast DNA type is of interest in the study of cytoplasmic male sterility, as well as in the study of various agronomic traits. For example, hybrids with cytoplasmic DNA W/γ and W/α were found to have high starch content compared to other types.

Previously, the plastomes of 15 potato cultivars of *S. tuberosum* grown in Russia were reconstructed and analyzed [1]. In this work, the sample of potato cultivars was expanded to 30.

Methods and Algorithms: Plastome reconstruction of an additional 15 potato cultivars of *S. tuberosum* based on short reads of Illumina sequencing technology was performed. Potato chloroplast DNA type was determined using *in silico* PCR and restriction endonuclease analysis. The reconstruction of potato phylogeny was performed in conjunction with previously collected plastomes as well as cultivars from other countries based on complete sequences of the plastomes [1].

Results: Genes encoding proteins, tRNAs and rRNAs were identified on the basis of *de novo* assembled plastomes. SNPs and InDels were identified in plastomes, microsatellites were quantitatively analyzed, and phylogeny was reconstructed. Potato chloroplast DNA types (W- and T-types) were identified based on *in silico* methods.

Conclusion: Our analysis provides the basis for further genetic studies of Russian potato cultivars, including the study of the plastome structure's relationship with the agronomic characteristics of plants.

Funding: The work was funded by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation, No. 075-15-2019-1662. This research was supported in part through computational resources of HPC facilities at collaborative center "Bioinformatics" ICG SB RAS.

References

1 Karetnikov D.I., Salina E.A., Kochetov A.V., Afonnikov D.A. Assembly and Analysis of Plastomes for 15 Potato Cultivars Grown in Russia. *Agronomy.* 2023;13:1454. doi 10.3390/agronomy13061454

doi 10.18699/bgrs2024-06-40

Analysis of genome structure and its variations in potato cultivars grown in Russia

Karetnikov D.I.[1, 2]*, Vasiliev G.V.[1, 2], Toshchakov S.V.[3], Shmakov N.A.[1, 2], Genaev M.A.[1, 2], Nesterov M.A.[1, 2], Ibragimova S.M.[1, 2], Rybakov D.A.[4], Gavrilenko T.A.[4], Salina E.A.[1, 2], Patrushev M.V.[3], Kochetov A.V.[1, 2], Afonnikov D.A.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Kurchatov Institute, Moscow, Russia*

[4] *N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia*

* *karetnikovmit@bionet.nsc.ru*

Key words: potato; *Solanum tuberosum* L.; genome structural variation; copy number variation; single nucleotide polymorphism; potato diversification

Motivation and Aim: *Solanum tuberosum* L. (potato) is the most important crop grown almost all over the world. The total production is more than 370 million tons per year and the number of cultivars is more than 4000. A peculiarity of potato genetics is the autotetraploid genome and high heterozygosity due to the predominantly vegetative mode of reproduction.

Whole-genome sequencing opens the way to study genome structure, identify single-nucleotide polymorphisms associated with different phenotypic traits, and reconstruct the pangenome at the level of a single species. One of the important features in genome analysis is copy number variation (CNV), defined as regions of the genome ranging in size from 1 kilobase to several megabases that exhibit different copy numbers in a population. At the functional level, CNV is associated with genes involved in stress responses and resistance to various factors.

Previously, 15 potato cultivars of *S. tuberosum* grown in Russia were sequenced [1]. Based on the sequenced genomes, the pangenome was reconstructed and potential CNVs associated with the genes were identified and analyzed. In this work, the sample of potatoes grown in Russia was expanded to 30.

Methods and Algorithms: DNA from 30 *S. tuberosum* cultivars was sequenced by paired-end short reads using the Illumina platform, which allowed DNA assembly at the exome level. Based on the short reads and DM1-3 v4.04 reference genome, CNVs in the genomes of the 30 cultivars were identified using the CNVpytor. In addition, we identified single nucleotide polymorphisms in all chromosomes, including extra chromosomes by GATK software package.

Results: Based on the collected genomes, we reconstructed the potato pangenome of *S. tuberosum*. CNV analysis revealed quantitative differences between cultivars grown in Russia and South American cultivars. Also, short reads made it possible to identify single nucleotide polymorphisms in all 30 genomes and conduct their further quantitative analysis.

Conclusion: Our findings lay a foundation for subsequent genetic research on Russian potato varieties, specifically focusing on how the structure of the genome and single nucleotide polymorphisms correlates with the plants' agronomic traits.
Funding: The work was funded by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation, No. 075-15-2019-1662. This research was supported in part through computational resources of HPC facilities at collaborative center “Bioinformatics” ICG SB RAS.

References

1. Karetnikov D.I., Vasiliev G.V., Toshchakov S.V. et al. Analysis of Genome Structure and Its Variations in Potato Cultivars Grown in Russia. *Int J Mol Sci*. 2023;24:5713. doi 10.3390/ijms24065713

doi 10.18699/bgrs2024-06-41

Functional characterization of genes with daily expression patterns in common wheat

Kiseleva A.A.[1, 2]*, Salina E.A.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *antkiseleva@bionet.nsc.ru*

Key words: common wheat; RNA-seq; daily expression pattern; circadian rhythms

Motivation and Aim: Plant circadian rhythms represent daily changes in the activity of various processes, which are based on changes in the levels of gene expression and protein synthesis. In wheat, some key components of plant circadian clock have been identified, but there is little data on the daily expression and interactions of these genes.
Methods and Algorithms: Common wheat cultivar Chinese Spring (CS) plants were cultivated for 21 days post-germination in controlled conditions with short days. Samples were harvested at 0, 3, 9, and 16-hour intervals over 24 hours, reflecting various stages of the light cycle. RNA was extracted and sequenced using a NextSeq 550 sequencer, resulting in over 420 million sequences deposited in the Sequence Read Archive NCBI database. Transcript mapping to the reference bread wheat genome and assembly were performed, followed by expression level assessment and normalization. Rhythmic gene expression analysis was conducted using JTK-CYCLE and Cosinor algorithms, with subsequent module separation using the R CemiTool package. Functional enrichment analysis was carried out using AgriGOv2.0 and g:Profiler, with visualization provided by the REViGO web server.
Results: In this work, for the first time, a comprehensive analysis of the daily transcriptome of common wheat was performed. There were 1421 genes with daily expression pattern divided into 5 modules according to their expression peak. The largest number of genes was expressed at the evening time point. GO enrichment analysis revealed biological processes in each module. The results indicated that plants prepare for sensing signals from the environment (predominantly light stimuli) at the end of the night before the light is turned on. During the day and in the evening, plants actively synthesize and accumulate different metabolites, including carbohydrates, lipids and a number of secondary metabolites. These processes are aimed at the active growth and development of plants. During the night period, all processes decrease; plants have a "rest". A comparison of daily expression patterns of genes of the central oscillator of common wheat with homologous genes of *Arabidopsis* demonstrated that wheat has homologs for most key circadian genes of *Arabidopsis*. Most of them were represented by three homoeologous genes. The expression patterns of core loop genes coincided, indicating a similar functional role of these genes in wheat. Although the expression patterns of some accessory genes were different compared to *Arabidopsis*, the whole mechanism of the central oscillator was uniform while changing the individual parts.
Conclusion: Thus, one can assume a high level of conservatism of the system of circadian clock in different plant species.
Funding: The study is supported RNA sequencing was funded by the RSF (Russian Science Foundation) project No. 21-76-30003.

doi 10.18699/bgrs2024-06-42

Genomic prediction in plants using machine learning

Kozlov K.*, Bankin M., Samsonova M.
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
** kozlov_kn@spbstu.ru*

Key words: bread wheat; rye; chickpea; genomic prediction; machine learning

Motivation and Aim: Today, genomic selection (GS) has become an established methodology in many plant breeding programs with the main application of reducing the length of breeding cycles [1]. Many genomic prediction (GP) methods have been proposed as there is no universal best method that can be used under all circumstances. The accuracy of GP depends on a series of factors including the quality and pre-processing of the phenotypic data, the platform used to obtain genomic information, the population mating design, the intrinsic genetic architecture of the trait, the genetic structure of the population, how the genotype-by-environment interaction is dealt with, and prediction method [2]. Nonparametric machine learning methods, e. g. Support Vector Machine, Random Forest, Gradient Boosting Machine provide tremendous flexibility to adapt to complicated associations between data and output [3]. Recently convolutional neural networks (CNN) have been intensively studied and applied in genome-based breeding [4]. These methods automatically identify latent patterns or features from data by sequentially stacking several layers. CNN are widely used in image classification [5]. Large genomics data can be encoded as artificial image objects (AIO) by considering individual genetic variants as pixels [6]. Consequently, AIO could be used by CNN on regression and classification tasks [7]. Here we used three datasets of bread wheat, rye and chickpea to compare the performance of several parametric and nonparametric methods in the regression task, i. e. in predicting phenotype from genomic data.

Methods and Algorithms: The rye and bread wheat datasets were generated by the Russian Bread consortium. Rye dataset consisted of 35208 SNPs, which were called in 451 samples phenotyped for 29 different morphological, productivity and biotic stress traits. The bread wheat dataset contained 16432 SNPs called in 960 samples, which were phenotyped for 7 traits. Several GP models were constructed for rye dataset namely BLUP, Bayessian regression with and without kinship matrix or covariance matrix, Extreme Gradient Boosting with and without two principal components. For bread wheat dataset GP models were build using BLUP, Bayessian regression with and without kinship matrix, as well as Extreme Gradient Boosting with two principal components and kinship matrix. Each model was constructed using 90 % of the dataset for training with the 4-fold cross-validation method. The rest of the samples comprising 10 % of the dataset were used to estimate the model accuracy by calculating the Pearson correlation between data and model prediction. All the models were trained for 30 iterations using various combinations of hyperparameters, with 10 different random partitions explored for each combination.

The combination of CNN for feature extraction and Extreme Gradient Boosting for regression was applied to model two main productivity traits of chickpea, 1000 seed weight (TSW) and number of seeds per plant (SNpP). Details on the phenotyping,

genotyping and subsequent analysis were presented in [8]. SNPs were encoded AIO and fed into CNN. Each convolutional kernel slides along the corresponding image channel and the results of processing are combined into one feature map. The size of the filter kernel of each convolutional layer was tuned using a small subset of the available experimental data and the kernel weights are the learning parameters of the model. Each convolutional layer is followed by a subsampling layer, which reduces the dimension of maps in order to enlarge features. Such filtering helps, among other things, to avoid overfitting. The formation of a new feature map is based on the max pooling operation, which is performed by selecting the maximum value from a subsample of a given size. The contribution of individual SNP to the model prediction was estimated by the feature attention map calculated as the weighted sum of the convolution layers of the network.

Results: Genomic prediction models were built for 15 rye traits. For some traits, e. g. Width of the third subflag leaf, several models resulted in accurate predictions, while for others (e. g., Weight of 1000 seeds) only one model was accurate. The maximum accuracy equal to 88 % was attained with the Extreme Gradient Boosting method for the Length of the third subflag leaf.

Predictive models were constructed for all wheat traits. The accuracy of the models for Brown rust, Yellow rust, Heading, Booting, Plant height and Yield exceeded 50 %, the best accuracy of 72.5 % was attained for the Yellow rust trait.

The accuracy of genomic prediction models for the TSW and SNpP traits of chickpea was 85 % and 84 %, respectively. Functional characterzation of 200 SNPs with the highest contribution to the TSW trait prediction showed that 91 of them are within known genes involved in general energy metabolism, photosynthesis, abiotic stress response, as well as in seed formation and quality. Likewise, 61 of 200 SNPs with the highest contribution to the SNpP trait prediction were within 50 known genes. 31 SNPs were found in the non-coding region of genes (introns) and 30 in the coding region (exons). Interestingly, 3 of these SNPs were found in gene encoding boron transporter 2. Boron plays essential roles in formation and stability of the plant cell wall through the cross-linking of the rhamnogalacturonan II [9].

Conclusion: The non-parametric methods yielded superior results than parametric methods across all traits and datasets. The best accuracy of 88 % was achieved for the Length of third subflag leaf on the rye dataset using the Extreme Gradient Boosting model. Accurate prediction models can be generated by using CNN in combination with AIO for feature extraction, however further studies are necessary to provide new valuable and unbiased information for assessing their methodological potential for GP.

Funding: The authors thank the Russian Bread Consortium, which provided access to datasets on rye and bread wheat. The study is supported by the Russian Science Fund (Grant No. 22-46-02004).

References

1. Meuwissen T.H.E., Hayes B.J., Goddard M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157:1819-1829. doi 10.1093/genetics/157.4.1819
2. de Los Campos G., Hickey J.M., Pong-Wong R., Daetwyler H.D., Calus M.P. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. 2013;193(2):327-345. doi 10.1534/genetics.112.143313
3. Montesinos-López O.A., Montesinos-López A., Mosqueda-Gonzalez B.A. et al. A zeroaltered Poisson random forest model for genomic-enabled prediction. G3 (Bethesda). 2021;11(2):jkaa057. doi 10.1093/g3journal/jkaa057
4. Montesinos-López O.A., Montesinos-López A., Pérez-Rodríguez P. et al. A review of deep learning applications for genomic selection. *BMC Genomics*. 2021;22(1):19. doi 10.1186/s12864-020-07319-x

5. Pook T., Freudenthal J., Korte A., Simianer H. Using Local Convolutional Neural Networks for Genomic Prediction. *Front Genet*. 2020;11:561497. doi 10.3389/fgene.2020.561497
6. Galli G., Sabadin F., Yassue R.M. et al. Automated Machine Learning: A Case Study of Genomic "Image-Based" Prediction in Maize Hybrids. *Front Plant Sci*. 2022;13:845524. doi 10.3389/fpls.2022.845524
7. Bavykina M. et al. Modeling of Flowering Time in Vigna radiata with Artificial Image Objects, Convolutional Neural Network and Random Forest. *Plants*. 2022;11:3327
8. Sokolkova A., Bulyntsev S.V., Chang P.L. et al. Genomic Analysis of Vavilov's Historic Chickpea Landraces Reveals Footprints of Environmental and Human Selection. *Int J Mol Sci.* 2020;21:3952
9. Saouros S., Mohan T.C., Cecchetti C. et al. Structural and functional insights into the mechanism of action of plant borate transporters. *Sci Rep*. 2021;11:12328. doi org/10.1038/s41598-021-91763-6

doi 10.18699/bgrs2024-06-43

Testing CAPS-markers designed on partially phased assembly of autotetraploid *Solanum tuberosum* genome

Larichev K.*, Sergeeva E., Karetnikov D., Afonnikov D.**, Salina E., Kochetov A.
Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** klarichev@bionet.nsc.ru; ** ada@bionet.nsc.ru*

Key words: *Solanum tuberosum*; genetic polymorphism; phased assembly; Illumina; CAPS markers

Motivation and Aim: As one of the key food crops, potato (*Solanum tuberosum* L.) is an attractive target for breeding. However, potato breeding is significantly complicated by its autotetraploid genome (2n = 4x = 48), as well as high heterozygosity and allelic diversity caused by clonal methods of reproduction. While marker assisted selection (MAS) can help alleviate some of the complications associated with tetraploid genome, taking the full advantage of MAS requires extensive knowledge of genetic polymorphisms that can be used as genetic markers. One of the most efficient ways to discover new genetic polymorphisms is by whole-genome sequencing. Recently, with the advent of long read sequencing and Hi-C, phased genome assembly, which provides the most comprehensive information regarding genetic polymorphisms, became viable even for polyploid species such as potato [1, 2]. However, such tools are quite expensive and not always available.

While the information about genomic variability acquired by short read sequencing may be incomplete (reads from different homologous chromosomes can be erroneously assembled into one contig), it can still be used to discover new genetic polymorphisms. Here we demonstrate how partially phased sequences assembled with Illumina short reads can be used to successfully discover novel genetic polymorphisms, even in organisms with complex genomes such as potato.

Methods and Algorithms: For our search of novel genetic polymorphisms we focused on 9 genes encoding the key enzymes of plant carbohydrate metabolism (ADP-glucose pyrophosphorylases *AGPase S3*, *AGPase S1*, beta-amylase *BAM1*, sucrose synthase *SuSy4*, starch phosphorylases *Pho1*, starch synthases *GBSS1*, *SS1*, *SS4*, Rubisco activase). Previously, using Illumina short reads, our colleagues assembled genomes of five Russian potato cultivars: Fritella, Krasavchik, Kolobok, Krepysh, Krasa Meschery. The coverage of assemblies on *S. tuberosum* DM1-3 R44 SolTub_3.0 genome assembly [3] ranged from 51 to 108 [4]. Reads belonging to selected genes were *de novo* assembled by SPAdes with default parameters and phased with freebayes and whatshap programmes. As a reference, we used gene sequences from DM1-3. The resulting partially phased gene sequences, as well as reference sequences and transcripts were aligned using MUSCLE algorithm in MEGA X. Conserved (no indels or substitutions) and polymorphic regions were annotated manually.

To verify SNPs predicted by assemblies, we used cleaved amplified polymorphic sequence (CAPS) assay. PCR primers for the assay were designed in Primer3 with the following considerations: (1) primers should anneal to conserved regions (preferably exons); (2) the length of PCR products should be 350-1100 bp; (3) each pair amplifies different parts of a gene; (4) amplification products should contain as least one SNP detectable using CAPS assay, i. e., one of the SNP alleles is a part of a restriction site.

Designed primers were used for CAPS assay of five Russian potato varieties (Fritella, Krasavchik, Kolobok, Krepysh, Krasa Meschery) in two replicas. Briefly, PCR products, amplified using cultivar genomic DNA as a template, were digested by an appropriate restriction enzyme (SibEnzyme, Russia) according to manufacturer's instructions. Digestion products were fractionated by agarose gel electrophoresis. Digestion patterns, corresponding to different SNP alleles, and the resulting cultivar allele distribution were discerned manually. PCR amplification was performed according to the following protocol: initial denaturation: 95 °C 5 min., denaturation 95 °C 15 sec., annealing 60 °C 15 sec., elongation 72 °C 30–90 sec. depending on the product length, final elongation 72 °C 5 min., 25 cycles in total.

Results: After *de novo* assembly and phasing, only two of the four haplotypes were resolved for each gene. For each cultivar, the resolved haplotypes demonstrated different SNP distribution. Using the assembled sequences, and taking into account SNP distribution in different cultivars, we designed 27 primer pairs amplifying various SNP containing regions of chosen genes. PCR products contained up to six SNPs, with three being the average. Observed amplicon length generally correlated to the predicted length in assembly. In some cases, deviations of 30-50 bp were observed, presumably caused by indels.

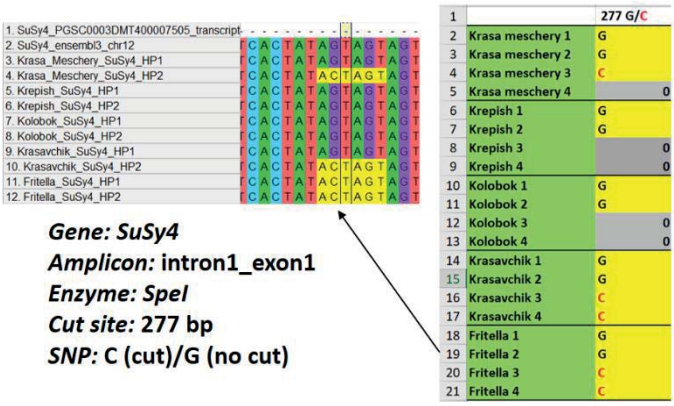


Fig. 1. Predicted haplotypes alignment in MEGA X and observed data obtained by CAPS assay

By using CAPS assay, we were able to detect more allelic variants (3–4) than was predicted by individual assemblies for most of the genes. For example, in the Fritella cultivar assembly, one of the SNPs in gene *SuSy4* was represented as 'C' allele in both haplotypes, however, with CAPS assay we additionally detected allele 'G' (Fig. 1). Thus, because of partial phasing, some SNPs were hidden in our assembly. It is worth noting that while the Fritella cultivar assembly did not show polymorphism in this site, it did in other cultivar's assemblies, which is why we designated this site as a SNP in the first place. In total, only *BAM1* gene had the same amount of detected and predicted alleles.

Conclusion: We demonstrated that partially phased assemblies can be used to discover novel CAPS markers for potato cultivars genotyping. Our results indicate that while some SNPs may be hidden due to partial phasing, and thus designated as 'non-polymorphic', using assemblies of several cultivars can help alleviate this problem by

providing additional information about these sites. In other words, combined, these assemblies provide more information about SNPs than individually, which should be taken into account when using partially phased assemblies as a means to find new genetic polymorphisms.

Funding: The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, in agreement with the Ministry of Education and Science of the Russian Federation, No. 075-15-2019-1662.

References

1. Sun H., Jiao W.-B., Krause K. et al. Chromosome-Scale and Haplotype-Resolved Genome Assembly of a Tetraploid Potato Cultivar. *Nat Genet*. 2022;54:342-348. doi 10.1038/s41588-022-01015-0
2. Hoopes G., Meng X., Hamilton J.P. et al. Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. *Mol Plant*. 2022;15(3):520-536. doi 10.1016/j.molp.2022.01.003
3. The Potato Genome Sequencing Consortium. Genome Sequence and Analysis of the Tuber Crop Potato. *Nature*. 2011;475:189-195. doi 10.1038/nature10158
4. Karetnikov D.I., Vasiliev G.V., Toshchakov S.V. et al. Analysis of Genome Structure and Its Variations in Potato Cultivars Grown in Russia. *Int J Mol Sci*. 2023;24(6):5713. doi 10.3390/ijms24065713

doi 10.18699/bgrs2024-06-44

Secreted In Xylem (SIX) genes across isolates of flax pathogenic ascomycete *Fusarium oxysporum* f. sp. *lini*

Logachev A.[1]*, Stanin V.[1], Bankin M.[1], Rozhmina T.[3], Kanapin A.[2], Samsonova A.[2], Samsonova M. [1]

[1] *Mathematical Biology and Bioinformatics Laboratory, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

[2] *Center for Computational Biology, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

[3] *Federal Research Center for Bast Fiber Crops, Torzhok, Russia*

** logachev_aa@spbstu.ru*

Key words: SIX genes; *Fusarium oxysporum*; virulence; flax; phylogeny; GFP; differential expression

Motivation and Aim: Secreted In Xylem (SIX) genes include 14 unrelated families of small cysteine-rich secreted proteins (also called effectors), originally found in xylem sap of plants severing from fusariosis, and occurred exclusively in pathogenic strains of *Fusarium oxysporum* [1]. SIX genes are used by fungus to interact with different molecules in different pathways of plant immunity to prevent immune response [2]. These 14 SIX genes families are grouped in different ways in various special forms and even within one form [3]. It was shown that deletion of some SIX genes leaded to reduced virulence on host plant, unlike our knowledge about structure and function of the most SIX genes are still remained elusive [4]. First, we aimed to investigate repertoire, structure and phylogeny of SIX genes in 13 isolates of *F. oxysporum* f. sp. *lini* (*Folini*) with different virulence to flax, which genomes published earlier [5–7]. Then, we tried to determine whether SIX genes taking part in progression of fusariosis infection on flax or not, and how they are influenced on fungal virulence and pathogenicity.

Methods and Algorithms: SIX genes in *Folini* genomes were detected by homology search with ncbi blast+ toolkit. introns and exact gene borders in the genomes were analyzed with MIT GENSCAN web server and FindORF tool [8]. ML phylogeny was created with IQ-TREE 1.6.12 wherefore the best fitted model was predicted for our dataset by implemented ModelFinder [9]. According to its prediction we chose Tamura–Nei model and then constructed ML tree with non-parametric UltraFast Bootstrap algorithm [10]. Genomic DNA of *Folini* isolates for PCR was extracted with DNeasy Plant Kit (QIAGEN), while total RNA was obtained with RNeasy Kit (QIAGEN). Fungal protoplast preparation and transformation with pCT74 vector performed according to [11] with some modifications.

Results: Firstly, we established a repertoire of SIX gene families in thirteen genomes of *Folini* by homology search with known sequences of SIX genes from other special forms. We found that eleven genomes had similar repertoire and consisted of *SIX1*, *SIX7*, *SIX10*, *SIX12* and *SIX13* families. We could not find any SIX genes in other two genomes (F365 and F482), that was a surprise. After that, we validated SIX genes repertoire by PCR analysis on presence/absence of their sequences in genomic DNA. The PCRs confirmed our genomic search.

As SIX genes cannot be predicted by Augustus, we aimed to obtain exact gene borders, gene sequence and coding sequence of each SIX gene copy found in genomes by

homology search. *SIX1* of *Folini* did not have any paralogous sequences and always presented as a single-copy gene in each studied genome. Six1 proteins had 277 a. a. length and was identical in all genomes except F418 and F456. Based on the differences in length and polymorphic loci within sequences between the genomes we assigned copies from chromosome 12 as *SIX7a*, and as *SIX7b* copies found on chromosome 15. *SIX10* gene in *Folini* genomes always situated on chromosome 12 and usually presented in one intact copy of 619 bp length. Structural analysis of *SIX10* reveals a single large conserved intron of 166 bp length. *Folini SIX12* orthologs can be grouped in two sequence variants of which *SIX12a* are present on both chromosomes 12 and 13, while *SIX12b* was found only on chromosome 13. *SIX13* gene sequence was the largest among other *Folini* SIX genes with estimated length 1275 bp, did not have paralogs in individual genomes, always occurred on chromosome 12, and had two large introns (168 bp and 160 bp). Another interesting finding was the presence of approximately one-half of *SIX8* gene sequence (493 bp of 1010 bp complete cds) on chromosome 13 in *Folini* genomes. Unfortunately, we could not find any correlation between SIX genes features typical for a given isolate and its virulence.

Interestingly, that genes *SIX7a*, *SIX12a* and *SIX13* formed a mini-cluster on chromosome 12 in MI39 genome, and the same cluster we found on chromosome 12 in F418 genome. Such mini-cluster were previously detected in Fol4287 isolate of *F. oxysporum* f. sp. *lycopersici* [12].

On the *SIX1* gene phylogenetic tree we observed clustering of *Folini* strains, surprisingly together with *F. oxysporum* f. sp. *cubense* (*Focub*) strain. Moreover, the clade which contains *Folini* also is settled by two separate subclades containing the rest of *Focub* and *F. oxysporum* f. sp. *canariensis* members that infect banana (*Musa* spp.) and canary palm (*Phoenix canariensis*), respectively. Phylogenetic inference on *SIX7* genes in *Folini* demonstrated differences between *SIX7a* and *SIX7b* sequence variants, since they occupy very distinct clades on the phylogenetic tree. On the *SIX10* gene phylogenetic tree we observed that all *Folini* strains were grouped into a separate clade distinct from other formae speciales. *SIX12* phylogeny clearly showed the divergence of *SIX12a* and *SIX12b* sequence variants which were placed into the separate clades in the phylogenetic tree. Interestingly, *SIX12b* clade also includes at least two members of *F. oxysporum* f. sp. *pisi*, that reaffirms the hypothesis of the horizontal gene transfer event between these formae speciales, which possibly took place in a recent past. Phylogenetic inference on *SIX13* gene demonstrated that *Folini* orthologs formed a distinct clade close to the clades consisted of *F. oxysporum* f. sp. *fragariae* and *F. oxysporum* f. sp. *niveum*, members of that causing severe on strawberry (Rosaceae) and watermelon (Cucurbitaceae), respectively.

To investigate whether SIX genes are involved in the virulence and could play a role during flax infestation process, we performed differential expression gene analysis from RNA-seq data in MI39 strain and validated it by RealTime-PCR. No expression of any studied SIX genes were found in mycelial culture. The most differentially expressed genes were *SIX1* and *SIX13*, and they also have the highest level of expression compare to other SIX genes. Expression in susceptible flax cultivar LM-98 began earlier than in resistant cultivar Atalanta: on third- and fifth-day post inoculation, respectively.

To further investigate role of SIX genes during plant tissue colonization we aimed to create mutant MI39 strain with deletion of *SIX13* gene, as one of the strongest differentially expressed SIX gene *in planta*. To obtain this mutant we firstly characterized stages of fusarium infectious in flax by creating GFP producing MI39

mutant to get easy microscopic observation of fungal structures and damaged plant cell. We transformed *Folini* MI39 wild type strain by pCT74 vector with PEG/$CaCl_2$ method and obtained transformant MI39_GFP2, which constantly expressed GFP under strong eukaryotic promoter.

Conclusion: The repertoire of SIX genes in *Folini* genomes is quite conserved and consists of five families: *SIX1*, *SIX7*, *SIX10*, *SIX12* and *SIX13*, with notable exception for two low-virulent strains F365 and F482. Each of *SIX1*, *SIX10* and *SIX13* genes usually present in one copy per genome, while *SIX7* and *SIX12* genes have paralogs and extra-copies in the individual genomes. In contrast to bulk of effectors that mostly placed in conservative genome part, SIX genes were exclusively found on variable chromosomes 12, 13 or 15. Differential expression analysis of SIX genes revealed that none of them was expressed in LCM. During infestation on flax all of detected SIX genes in MI39 genome have expressed *in planta*, but with different level of expression. We created MI39_GFP2 mutant which constantly expressed GFP and described stages of fusarium infection during flax colonization.

Funding: The study is supported by RSF grant (No. 23-16-00037).

References

1. Rep M., Dekker H.L., Vossen J.H., de Boer A.D., Houterman P.M., Speijer D., … Cornelissen B.J. Mass spectrometric identification of isoforms of PR proteins in xylem sap of fungus-infected tomato. *Plant Physiol.* 2002;130(2):904-917
2. Rep M., Van Der Does H.C., Meijer M., Van Wijk R., Houterman P.M., Dekker H.L., … Cornelissen B.J. A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for I-3-mediated resistance in tomato. *Mol. Microbiol.* 2004;53(5):1373-1383
3. Taylor A., Vágány V., Jackson A.C., Harrison R.J., Rainoni A., Clarkson J.P. Identification of pathogenicity-related genes in *Fusarium oxysporum* f. sp. *cepae*. *Mol. Plant Pathol.* 2016;17(7):1032-1047
4. Jenkins S., Taylor A., Jackson A.C., Armitage A.D., Bates H.J., Mead A., … Clarkson J.P. Identification and expression of secreted in xylem pathogenicity genes in *Fusarium oxysporum* f. sp. *pisi*. *Front. Microbiol.* 2021;12:593140
5. Kanapin A., Samsonova A., Rozhmina T., Bankin M., Logachev A., Samsonova M. The genome sequence of five highly pathogenic isolates of *Fusarium oxysporum* f. sp. *lini*. *Mol. Plant-Microbe Interact.* 2020;33(9):1112-1115
6. Dvorianinova E.M., Pushkova E.N., Novakovskiy R.O., Povkhova L.V., Bolsheva N.L., Kudryavtseva L.P., … Dmitriev A.A. Nanopore and Illumina genome sequencing of *Fusarium oxysporum* f. sp. *lini* strains of different virulence. *Front. Genet.* 2021;12:662928
7. Kanapin A.A., Samsonova A.A., Bankin M.P., Logachev A.A., Rozhmina T.A., Samsonova M.G. Assembly of the Genomes of Three Weakly Virulent *Fusarium oxysporum* f. sp. *lini* Strains. *Biophysics*. 2022;67(2):180-182
8. Burge C.B., Karlin S. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* 1998;8(3):346-354
9. Nguyen L.T., Schmidt H.A., Von Haeseler A., Minh B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 2015;32(1):268-274
10. Hoang D.T., Chernomor O., Von Haeseler A., Minh B.Q., Vinh L.S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 2018;35(2):518-522
11. Oren L., Ezrati S., Cohen D., Sharon A. Early events in the *Fusarium verticillioides*-maize interaction characterized by using a green fluorescent protein-expressing transgenic isolate. *Appl. Environ. Microbiol.* 2003;69(3):1695-1701
12. Schmidt S.M., Houterman P.M., Schreiver I., Ma L., Amyotte S., Chellappan B., … Rep M. MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics*. 2013;14:1-21

doi 10.18699/bgrs2024-06-45

Characterization of two chloroplast genomes in Cornaceae: *Cornus sanguinea* and *Cornus sericea*

Nikonorova E.[1]*, Khapilina O.[2], Sereda A.[1], Bondarev S.[1], Krol' T.[1]

[1] *All-Russian Research Institute of Medicinal and Aromatic Plants (VILAR), Moscow, Russia*

[2] *National Center for Biotechnology, Astana, Kazakhstan*

* *gatiatulinaer@gmail.com*

Key words: chloroplast genomes; *Cornus*; phylogenetics; genomic diversity

Motivation and Aim: The Cornaceae family encompassing over a hundred species, represents a significant group in the order *Cornales* [1]. The genus *Cornus*, particularly noted for its ornamental and medicinal value, comprises species whose phytochemical-rich fruits contribute to traditional medicine, treating a range of pathological conditions from diabetes to rheumatic disorders [2]. Despite extensive studies, the phylogenetic relationships within the family remain complex and controversial [3]. This study aims to clarify these relationships by analyzing the complete chloroplast (cp) genomes of *Cornus sanguinea* and *Cornus sericea*, focusing on their structure and gene content.

Methods and Algorithms: Mature leaves of *C. sanguinea* and *C. sericea* were collected in July at the Botanical Garden of VILAR, followed by chloroplast isolation using high ionic strength solutions and DNA extraction. Sequencing libraries were generated using Illumina DNA Prep, (M) Tagmentation kit and sequenced using the MiSeq Illumina platform. Read quality was assessed via fastqc, and adapter trimming and filtering of low-quality reads were performed using BBduk within Geneious. The plastome assembly was performed using the GetOrganelle toolkit and annotations were carried out with GeSeq, PGA, and CPGAVAS2 tools using *C. capitata* (NC_084212.1) as the reference genome, with subsequent manual review and annotation correction [4]. Simple sequence repeats (SSRs) were determined using the MISA software. For the phylogenetic study, plastomes within the Cornaceae, Hydrangeaceae, Nyssaceae, Garryaceae, Curtisiaceae, Grubbiaceae families, and the *Arabidopsis thaliana* plastome were downloaded from NSBI. Multiple sequence alignment was performed using MAFFT followed by trimming using TrimAl. The phylogenetic analysis was carried out using the neighbor-joining method with 500 bootstrap replicates and the Tamura–Nei model.

Results: The analysis revealed that both *C. sanguinea* and *C. sericea* possess a typical quadripartite structure of the chloroplast genome, with a slight variation in the size of the Large Single Copy (LSC) and Small Single Copy (SSC) regions compared to other *Cornus* species. The cp genomes had typical size being 158,244 and 158,663 bp for *C. sericea* and *C. sanguinea*, respectively. The cp genomes of *C. sericea* and *C. sanguinea* were found to contain 131 genes each, including 86 protein-coding genes, 37 tRNA genes, and 8 rRNA genes (Table 1, the number of unique genes is shown in brackets).

Table 1. The basic chloroplast genome characteristics of *C. sericea* and *C. sanguinea*

Species		*C. sericea*	*C. sanguinea*
Genome size (bp)		158,244	158,663
GC content (%)		37.9	37.8
SSC	Genome size (bp)	18,711	18,644
	GC content (%)	32.0	31.9
LSC	Genome size (bp)	87,459	87,925
	GC content (%)	36.0	36.0
IRs	Genome size (bp)	26,037	26,047
	GC content (%)	43.1	43.0
CDS		86 (79)	86 (79)
gene		131 (113)	131 (113)
rRNA		8 (4)	8 (4)
IRs		2	2
tRNA		37 (30)	37 (30)

These results in general are align with some of previous studies [5], but slightly differenced from genes number in *C. bretschneideri*, which was higher (132 vs 131) [6], and *C. alba*, where the number of genes and tRNA was 132 and 38, respectively [7]. At the same time, an only 113 unique genes, of which 79 were protein-coding genes, 30 tRNA genes, and 4 rRNA genes were identified in cp genome of *C. sunhangii* [8]. Phylogenetic analysis placed *C. sanguinea* and *C. sericea* within a broader clade of the Cornaceae family, reflecting their close genetic relationship with other species in the family, which is consistent with existing studies [5, 9, 10]. In present study the differences in the inverted repeats and their junctions with SSC and LSC in 8 representative *Cornus* species were found. A total of 53 SSRs were revealed for each species, which was higher than that in reference genome *C. capitata*, and there was a complete absence of tetra-, penta-, and hexa-nucleotide repeats in the sequenced species. These results were differed from those obtained by Guan et al. [5], where the only 25–31 SSRs were identified in 10 taxa of *Cornus* subg. *Syncarpea*.

Conclusion: This study provided detailed insights into the chloroplast genomes of *C. sanguinea* and *C. sericea*, underscoring the genetic stability and minimal variation across the Cornaceae family. The findings support the hypothesis of conserved genome structures within the family, which could be pivotal for future phylogenetic and evolutionary studies. The detailed characterization of these genomes enhances our understanding of the adaptive strategies of these species, with implications for conservation biology and the use of genetic resources in breeding and cultivation of these species.

Funding: The study was carried out in accordance with the State task on the topic FGUU-2024-0001.

References

1. Simpson M.G. Diversity and Classification of Flowering Plants: Eudicots. In: Plant Systematics. Elsevier, 2010;275-448. doi 10.1016/B978-0-12-374380-0.50008-7
2. Dinda B., Kyriakopoulos A.M., Dinda S. et al. Cornus mas L. (cornelian cherry), an important European and Asian traditional food and medicine: Ethnomedicine, phytochemistry and pharmacology for its commercial utilization in drug industry. *J Ethnopharmacol*. 2016;193:670-690. doi 10.1016/j.jep.2016.09.042

3. Du Z.Y., Jenny Xiang Q.Y., Cheng J. et al. An updated phylogeny, biogeography, and PhyloCode-based classification of Cornaceae based on three sets of genomic data. *Am J Bot*. 2023;110:e16116. doi 10.1002/ajb2.16116
4. Amiryousefi A., Hyvönen J., Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics*. 2018;34:3030-3031. doi 10.1093/bioinformatics/bty220
5. Guan B., Wen J., Guo H., Liu Y. Comparative and phylogenetic analyses based on the complete chloroplast genome of Cornus subg. Syncarpea (Cornaceae) species. *Front Plant Sci*. 2024;15:1306196. doi 10.3389/fpls.2024.1306196
6. Li X., Ma Q., Zhou H., Yang Y., Li H., Wang J. Characterization of the complete chloroplast genome of *Cornus bretschneideri* (cornaceae). *Mitochondrial DNA B Resour*. 2020;5(1):543-544. doi 10.1080/23802359.2019.1710281
7. Yuan W., He S., Zhang S., Chang D., He Y. The complete chloroplast genome sequence of *Cornus Alba* L. (Cornaceae). *Mitochondrial DNA B Resour*. 2021;6(7):1997-1998. doi 10.1080/23802359.2021.1938727
8. Lv Z.-Y., Huang X.-H., Luo J., Zhang X., Deng T., Li Z.-M. The complete chloroplast genome sequence of *Cornus sunhangii* (Cornaceae). *Mitochondrial DNA B Resour*. 2019;4(2):3242-3243. doi 10.1080/23802359.2019.1669090
9. Keir K.R., Bemmels J.B., Aitken S.N. Low genetic diversity, moderate local adaptation, and phylogeographic insights in Cornus nuttallii (Cornaceae). *Am J Bot.* 2011;98:1327-1336. doi 10.3732/ajb.1000466
10. Fu C.N., Li H.T., Milne R. et al. Comparative analyses of plastid genomes from fourteen Cornales species: inferences for phylogenetic relationships and genome evolution. *BMC Genomics*. 2017;18(1):956. doi 10.1186/s12864-017-4319-9

doi 10.18699/bgrs2024-06-46

FindTFnet: a tool for reconstruction of Arabidopsis transcription factor networks from transcriptome data

Omelyanchuk N.A.[1]*, Lavrekha V.V.[1, 2], Bogomolov A.G.[1], Zemlyanskaya E.V.[1, 2]

[1] *Department of Systems Biology, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Department of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

* *nadya@bionet.nsc.ru*

Key words: transcriptome; transcription factor; gene network

Motivation and Aim: Transcription factors (TFs) are proteins that bind to regulatory DNA sequences mainly in gene promoters and regulate transcription of these genes [1]. TFs directly regulating transcription of genes encoding other TFs compose transcription factor regulatory networks (TFRNs) [2]. In actions of exogenous or endogenous factors, TFRNs are intermediate between signaling pathway, which they activate and the processes they affect. Nevertheless, TFRNs are the less studied elements in these biological chains. Here we present FindTFnet – a tool for the reconstruction of TFRNs in *Arabidopsis thaliana* from transcriptome data on differentially expressed genes (DEGs) and a list of their upstream regulators predicted with the CisCross web service [3]. We illustrate the efficiency of FindTFnet by its application for the reconstruction of mechanisms underlying auxin response.

Methods and Algorithms: The lists of DEGs down- and upregulated by auxin (dDEGs and uDEGs, respectively) were compiled based on the analysis of microarray data on auxin-induced changes in the *A. thaliana* root [4]. First, we identified the lists of TFs, for which DAP-seq peaks were enriched in 1000 bp-long upstream regions of DEGs. According to the enrichment status, TFs were subdivided into three following groups: overrepresented in dDEGs, in uDEGs, and in both dDEGs and uDEGs (Table 1). The corresponding TF-coding genes can be dDEG, uDEG or not transcriptionally regulated (NTRs).

Table 1. The scheme for determining TF functions in a TFRN reconstruction

	TF is enriched in dDEG promoters	TF is enriched in uDEG promoters	TF is enriched in both dDEG and uDEG promoters
TF gene is dDEG	Downregulated activator – DA	Downregulated suppressor – DS	Downregulated Regulator (DR)
TF gene is uDEG	Upregulated suppressor – US	Upregulated activator – UA	Upregulated Regulator (UR)
TF gene is not DEG	NTR	NTR	NTR

DRs, URs and NTRs are out of scope of this research. If a TF is encoded by a dDEG, it is an auxin downregulated regulator. If this TF's binding loci are enriched in dDEG promoters, it should be a Downregulated Activator (DA), if they are enriched in uDEG promoters, this TF should be a Downregulated Suppressor (DS) (Table 1). Similarly, the functions in the TFRN are defined for TFs encoded by auxin-activated genes: enrichment of TF binding loci in dDEGs or uDEGs testifies that it should be an Upregulated

Suppressor (US) or an Upregulated Activator (UA), respectively. It is clear that downregulated TFs are active before auxin treatment, and since the amount of DAs and DSes is decreased upon auxin treatment, their targets are downregulated and upregulated passively. In this downregulated subnetwork (D-subnetwork) DAs activate other DAs and DSes, and DSes suppress UAs and USes. In opposite, auxin application switches on USes and UAs (U-subnetwork). USes actively suppresses DAs and DSes, and UAs actively activate other UAs and USes.

Results: FindTFnet is available as a part of CisCross web service [3]. Its application for the analysis of auxin-sensitive DEGs identified 60 TFs, which mediate response to auxin: 23 DAs, two DSes, eight USes and five UAs. Of them, 54 TFs constitute a connected transcriptional network (Fig. 1). The rest of TFs, which do not participate in any "TF regulator–TF target" pair, include one DA (LCL1), one DS (RAP2.12), two UAs (MYB3R1 and DEL2), and two USes (HB18 and NAM).

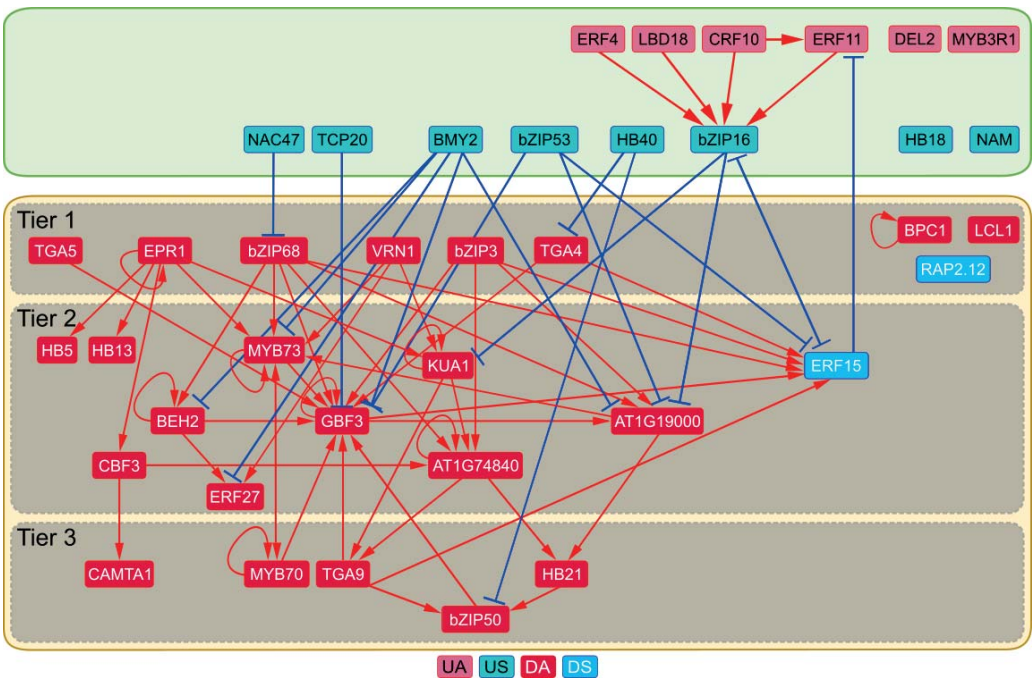


Fig. 1. Auxin depended TFRN in Arabidopsis roots inferred by FindTFnet. Three tiers of the downregulated subnetwork (D-subnetwork) denoted by grey boxes are encircled by yellow contours, an activated subnetwork (U-subnetwork) is enclosed in a green box

Of 60 TFRN TFs, the role in auxin response was previously demonstrated only for 12 TFs (25 %), namely, DAs: CAMTA1 [5]; HB5 [6], KUA [7], MYB70 [8], MYB73 [9], DS: RAP2.12 [10]; USes: BMY2 [11], HB40 [12], TCP20 [13]; UAs CRF10 [14], DEL2 [15], LBD18 [16].

Conclusion: FindTFnet (https://plamorph.sysbio.ru/ciscross/FindTFnet_index.html) reconstructs TF regulatory networks guiding gene expression changes in response to endogenous or environmental factors. The analysis of auxin-sensitive DEGs with FindTFnet allowed inferring the TFRN acting in auxin response, in which only for one fifth of TFs this role was previously recognized.

Funding: The study was supported by the RSF grant No. 20-14-00140.

References

1. Cramer P. Organization and regulation of gene transcription. *Nature*. 2019;573(7772):45-54. doi 10.1038/s41586-019-1517-4
2. Brent M.R. Past roadblocks and new opportunities in transcription factor network mapping. *Trends Genet*. 2016;32(11):736-750. doi 10.1016/j.tig.2016.08.009
3. Lavrekha V.V., Levitsky V., Tsukanov A.V., Bogomolov A.G., Grigorovich D.A., Omelyanchuk N., Zemlyanskaya E.V., Mironova V. CisCross: A gene list enrichment analysis to predict upstream regulators in *Arabidopsis thaliana*. *Front Plant Sci*. 2022;13:942710. doi 10.3389/fpls.2022.942710
4. De Rybel B., Audenaert D., Xuan W. et al. A role for the root cap in root branching revealed by the non-auxin probe naxillin. *Nat Chem Biol*. 2012;8(9):798-805. doi 10.1038/nchembio.1044
5. Galon Y., Snir O., Fromm H. How calmodulin binding transcription activators (CAMTAs) mediate auxin responses. *Plant Signaling Behav*. 2010;5(10):1311-1314. doi 10.1007/s00425-010-1153-6
6. De Smet I., Lau S., Ehrismann J.S. et al. Transcriptional repression of BODENLOS by HD-ZIP transcription factor HB5 in Arabidopsis thaliana. *J Exp Bot*. 2013;64(10):3009-3019. doi 10.1093/jxb/ert137
7. Moreno J.E., Romani F., Chan R.L. Arabidopsis thaliana homeodomain-leucine zipper type I transcription factors contribute to control leaf venation patterning. *Plant Signaling Behav*. 2018;13(3):577-590. doi 10.1080/15592324.2018.1448334
8. Huang C.K., Lo P.C., Huang L.F. et al. A single-repeat MYB transcription repressor, MYBH, participates in regulation of leaf senescence in Arabidopsis. *Plant Mol Biol*. 2015;88:269-286. doi 10.1007/s11103-015-0321-2
9. Wan J., Wang R., Zhang P., Sun L., Ju Q., Huang H., Lu S., Tran L., Xu J. MYB70 modulates seed germination and root system development in Arabidopsis. *iScience*. 2021;24(11):103228. doi 10.1016/j.isci.2021.103228
10. Yang Y., Zhang L., Chen P., Liang T., Li X., Liu H. UV-B photoreceptor UVR8 interacts with MYB73/MYB77 to regulate auxin responses and lateral root development. *EMBO J*. 2020;39(2):e101928. doi 10.15252/embj.2019101928
11. Shukla V., Lombardi L., Iacopino S., Pencik A., Novak O., Perata P., Giuntoli B., Licausi F. Endogenous hypoxia in lateral root primordia controls root architecture by antagonizing auxin signaling in Arabidopsis. *Mol Plant*. 2019;12(4):538-551. doi 10.1016/j.molp.2019.01.007
12. Reinhold H., Soyk S., Šimková K. et al. β-Amylase – Like proteins function as transcription factors in Arabidopsis, controlling shoot growth and development. *Plant Cell*. 2011;23(4):1391-1403. doi 10.1105/tpc.110.081950
13. Mora C.C., Perotti M.F., González-Grandío E., Ribone P.A., Cubas P., Chan R.L. AtHB40 modulates primary root length and gravitropism involving CYCLINB and auxin transporters. *Plant Sci*. 2022;324:111421. doi 10.1016/j.plantsci.2022.111421
14. Kong Q., Low P.M., Lim A.R., Yang Y., Yuan L., Ma W. Functional antagonism of WRI1 and TCP20 modulates GH3.3 expression to maintain auxin homeostasis in roots. *Plants*. 2022;11(3):454. doi 10.3390/plants11030454
15. Truskina J., Han J., Chrysanthou E. et al. A network of transcriptional repressors modulates auxin responses. *Nature*. 2021;589(7840):116-119. doi 10.1038/s41586-020-2940-2
16. Sozzani R., Maggio C., Giordo R. et al. The E2FD/DEL2 factor is a component of a regulatory network controlling cell proliferation and development in Arabidopsis. *Plant Mol Biol*. 2010;72:381-395. doi 10.1007/s11103-009-9577-8
17. Goh T., Toyokura K., Yamaguchi N. et al. Lateral root initiation requires the sequential induction of transcription factors LBD16 and PUCHI in Arabidopsis thaliana. *New Phytol*. 2019;224:749-760. doi 10.1111/nph.16065

doi 10.18699/bgrs2024-06-47

Formation of a pool of different-sized virus-specific short RNAs in *Solanum tuberosum*

Samarskaya V.[1]*, Spechenkova N.[1], Kalinina N.[1, 2], Taliansky M.[1]

[1] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

[2] *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

** viktoriya.samarskaya2012@yandex.ru*

Key words: RNAi; potato virus Y; small RNAs; *Solanum tuberosum*

Motivation and Aim: Potato plants are an important agricultural crop. However, potatoes can be affected by various diseases, including viral ones. RNA-based technologies, including external application of double-stranded RNA (dsRNA), can be used to protect crops from pests. While it is generally accepted that the mechanism of dsRNA-mediated antiviral RNA silencing reflects natural RNA interference (RNAi) defense against RNA viruses, there is little direct evidence to support this notion.

Methods and Algorithms: Potato virus Y strain NTN (PVY) was propagated on potato plants of the Manhattan variety. For the experiments, we used two-week-old Indigo potato plants grown in soil from seedlings propagated in vitro. For each experimental condition, potato plants were treated with dsRNA solution or buffer, and the next day they were inoculated with buffer or PVY and allowed to grow in a controlled environment chamber. The photoperiod of which was 16/8 hours day/night, relative humidity 40 % and light flux density 250 μmol m-2 s-1. Leaf samples were taken 7 days after the plants were treated with the potato virus.

Small RNA was prepared for sequencing using the NEXTFlex Small RNA-Seq v3 kit (Bio Scientific, Gymea, Australia) which were then sequenced using the Illumina NovaSeq 6000 with a single-read length of 50 nucleotides. After preparing the reads, reads were mapped to the PVY-NTN genome, with zero mismatches. Mapping results were visualized. To analyze viral siRNAs, the reference sequences of PVY-NTN genomes were used to map 18–30 nt reads from each library. The sorted sRNAs were then counted by size, polarity, and 5′-terminal nucleotide identity (5′A, 5′C, 5′G and 5′U).

Results: High-throughput sRNA sequencing was conducted to analyze the sRNA populations resulting from the application of dsRNApvy and the virus. Three treatments were examined: (i) PVY, (ii) dsRNApvy, and (iii) PVY + dsRNApvy in both treated and untreated leaves. The analysis revealed that PVY-infected plants contained virus-specific siRNA species of 21 nt and 22 nt in size, with 21 nt species being predominant, consistent with sRNA profiles induced by other RNA-containing plant viruses. In contrast, dsRNA-treated leaves of PVY-uninfected plants showed a range of sRNAs in length from 18–30 nt, suggesting a non-canonical origin. However, all samples exclusively matched the PVY genome segment used for dsRNA design, indicating a sequence-specific origin for the non-canonical sRNAs. Investigation into systemic movement of these non-canonical sRNAs showed that dsRNA applied to bottom leaves could move and accumulate in upper untreated leaves. In plants with deficient systemic movement, sRNA accumulation in untreated leaves was minimal. These findings suggest

that systemically moving dsRNA can be processed into non-canonical sRNAs in upper untreated leaves similar to treated leaves.

Upon aligning the mRNA with the PVY-NTN genome sequence, it was observed that in dsRNApvy-treated leaves of uninfected plants, the number of reads with sequences complementary to the viral genome (referred to as antisense mRNAs) exceeded the number of opposite strand-specific reads (sense mRNAs) across all mRNA size classes (18–30 nt). A similar pattern was noted in untreated leaves of dsRNApvy-treated uninfected plants. In PVY-infected plants, both strands of the entire PVY genome contributed to virus-induced mRNAs detected in inoculated and non-inoculated (systemically infected) leaves, albeit with a bias towards the sense strand.

In addition to mRNA size, the 5'-terminal nucleotide is crucial in determining the selective loading of mRNA into specific AGOs. To assess the potential number of RISC-producing mRNAs that could actively suppress PVY, the relative abundance of four different 5′ nucleotide identities in antisense mRNAs was examined. In PVY infection, two major classes of antisense mRNAs (21 nt and 22 nt) were enriched in 5'U (36–40 %) and 5'A (28–29 %), followed by 5'C (20 %) and 5'G (21 %) in both inoculated and systemically infected leaves, indicating their association with AGO1-, AGO2-, and AGO5-like proteins, respectively.

Conclusion: Thus, the presented results from high-throughput sequencing (HTS) analysis of virus-specific small non-coding RNAs (mRNAs) may help elucidate the precise mechanisms induced by external dsRNAs.

Funding: The study is supported by the Russian Science Foundation grant number 23-74-30003.

References

1. Ding S.-W. RNA-Based Antiviral Immunity. *Nat Rev Immunol.* 2010;10:632-644. doi 10.1038/nri2824
2. Baulcombe D.C. The Role of Viruses in Identifying and Analyzing RNA Silencing. *Annu Rev Virol.* 2022;9:353-373. doi 10.1146/annurev-virology-091919-064218
3. Qi Y., Denli A.M., Hannon G.J. Biochemical Specialization within Arabidopsis RNA Silencing Pathways. *Mol Cell.* 2005;19:421-428. doi 10.1016/j.molcel.2005.06.014
4. Samarskaya V.O., Spechenkova N., Ilina I., Suprunova T.P., Kalinina N.O., Love A.J., Taliansky M.E. A Non-Canonical Pathway Induced by Externally Applied Virus-Specific dsRNA in Potato Plants. *Int J Mol Sci*. 2023;24(21):15769. doi 10.3390/ijms242115769

doi 10.18699/bgrs2024-06-48

Fusarium pathogenesis: mechanisms of infection and disease progression in flax

Samsonova M.,[1]*, Kanapin A.[1], Rozhmina T.[2]
[1] *Peter the Great Polytechnic University, St. Petersburg, Russia*
[2] *Federal Research Center for Bast Fiber Crops, Torzhok, Russia*
* *m.g.samsonova@gmail.com*

Key words: flax; Fusarium wilt; pangenome; virulence; resistance

Motivation and Aim: Fusarium wilt is one of the most economically damaging flax diseases caused by a specific type of *Fusarium oxysporum* Schlecht., a fungus in the Ascomycota phylum [1]. This special form (i. e., forma specialis, f. sp.) of *Fusarium oxysporum* is known to infect only flax [2]. The mechanisms of flax resistance to Fusarium wilt have never been fully understood, although resistance to the disease was developed by selection. Here we applied the analysis of genomics data, as well as innovative breeding methods (GWAS and bulk segregant analysis) to (1) perform chromosme-level assembly of the fungus genome and reveal its complex multi-compartmentalized organization, (2) construct the pangenome to assess the genomic diversity of *Fusarium oxysporum* f. sp. *lini* (Folini) strains and (3) map several resistance genes identified previously by classical genetics methods to reference flax genome.
Methods and Algorithms: Strain virulence was assessed under greenhouse conditions using reference cultivars with contrasting susceptibility to Fusarium wilt. LRS, Illumina and Hi-C technologies were used to sequence and assemble the genomes of 36 *Folini* strains with different virulence status. Protein orthogroups were identified with silix software (version 1.3.0). Candidate secreted proteins were identified in three steps as described in [4]. GWAS on flax collection of 306 lines was conducted with GAPIT [5] and 3VmrMLM [6] packages.
Results: The size of *Folini* genome is about 65 MBp, it contains 14,000 gene models. Most *Folini* strains have 20 chromosomes. There are at least 3 partitions in *Folini* genome, which differ in conservation, the density of genes, SNPs and repeats, speed of evolution. The pangenome of 36 *Folini* isolates is open. The *Folini* pansecterome, encompassing secreted proteins critical for fungal pathogenicity, primarily consists of three functional classes: effector proteins, CAZYmes, and proteases. Each functional class within the pansecterome was meticulously annotated and characterized with respect to pangenome category distribution, PFAM domain frequency, and strain virulence assessment. Regions in plant genome associated with resistance were identified on 1, 8, 11 and 13 chromosomes.
Conclusion: Our results provide novel insights into plant-pathogen interactions and pinpoint potential candidate genes for further in-depth studies.
Funding: The study is supported by the RScF grant 23-16-00037.

References

1. Gordon T.R. Fusarium oxysporum and the Fusarium Wilt Syndrome. *Annu Rev Phytopathol.* 2017;55:23-39. doi 10.1146/annurev-phyto-080615-095919

2. Baayen R.P., O'Donnell K., Bonants P.J.M. et al. Gene Genealogies and AFLP Analyses in the Fusarium oxysporum Complex Identify Monophyletic and Nonmonophyletic Formae Speciales Causing Wilt and Rot Disease. *Phytopathology*. 2000;90:891-900. doi 10.1094/phyto.2000.90.8.891
3. Miele V., Penel S., Duret L. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics*. 2011;12:116. doi 10.1186/1471-2105-12-116
4. Samsonova A., Kanapin A., Bankin M. et al. A Genomic Blueprint of Flax Fungal Parasite Fusarium oxysporum f. sp. lini. *Int J Mol Sci*. 2021;22:2665. doi 10.3390/ijms22052665
5. Wang J., Zhang Z. GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinformatics*. 2021;19(4):629-640. doi 10.1016/j.gpb.2021.08.005
6. Li M., Zhang Y.-W., Xiang Y., Liu M.-H., Zhang Y.-M. IIIVmrMLM: The R and C++ Tools Associated with 3VmrMLM, a Comprehensive GWAS Method for Dissecting Quantitative Traits. *Mol Plant*. 2022;15:1251-1253. doi 10.1016/j.molp.2022.06.002

doi 10.18699/bgrs2024-06-49

Changes in the transcription activity of the *Ago1*, *Ago2* and *Dcl2* genes in various wheat species under the influence of infection with the causative agent of Septoria

Shein M.*, Burkhanova G., Veselova S., Maksimov I.
Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
** mikeshenoda@yandex.ru*

Key words: RNA interference; Ago and Dcl genes; influence of phytopathogen on the activity of RNAi genes

Motivation and Aim: The search for effective and environmentally friendly approaches to protecting cultivated plants from pathogens is one of the most pressing areas of plant science, since pathogens often lead to significant losses in crop yields. The discovery and explanation of the phenomenon of RNA interference (RNAi), which is involved in the regulation of the activity of various genes in both the host plant and the pathogen [1], made it possible to evaluate the possibility of its use for practical purposes. In this work, we analyzed the activity of the Ago1, Ago2 and DCl2 genes of the RNAi system in a number of representatives of the genus *Triticum* L.: *T. timopheevii*, *T. monococcum* and the *T. aestivum* variety "Salavat Yulaev" during infection with the pathogenic fungus *S. nodorum* Berk.
Methods and Algorithms: Experiments were carried out on wheat plants of the genus *Triticum* L.: *T. timopheevii* k-58666, *T. monococcum* k-39471 and *T. aestivum* of the moderately susceptible variety "Salavat Yulaev". The experiments were carried out in laboratory conditions, where the plants were grown at room temperature (20–22 °C) in a light area with a 16-hour light period. Before the experiments, wheat seeds were sterilized with potassium permanganate solution and washed several times with sterile water. Next, the seeds were germinated in enamel cuvettes on damp filter paper. The dishes and paper were autoclaved at a temperature of 150–180 °C for 1.5–2 hours. Fully expanded first leaves of 7-day-old seedlings were cut off and placed in Petri dishes on damp cotton wool with the addition of benzimidazole (40 mg/l). In the experiment, we used a highly virulent strain of the phytopathogenic fungus *T. aestivum* against common wheat, *Stagonospora nodorum* – SnB, from the collection of the Institute of Biogeochemistry, UFRC RAS. Infection of seedlings with pycnospores of the fungus *S. nodorum* was carried out by applying a spore suspension at a concentration of 105 spores/ml to a leaf with a micropipette with a volume of 4–5 μl per leaf. Isolation of total RNA to assess the level of transcripts of the studied genes was carried out using the Lyra reagent (Biolabmix, Russia), according to the protocol of the supplier. To synthesize cDNA, a reverse transcription reaction was performed using M-MuLV reverse transcriptase (Synthol, Russia). The transcript copy number for each gene under study was determined by quantitative real-time PCR on a CFX Connect Real-Time System device (Bio-Rad, USA) using the intercalating dye EVA-Green (Synthol, Russia) c using the software "Bio-Rad CFX Maestro 1.1 Version: 4.1.2433.1219" (Bio-Rad, USA) relative housekeeping reference gene: TaRLI (AY059462). Measurements

of transcription activity of the studied genes were carried out in 3 biological and 3 analytical replicates. Statistical processing of real-time PCR data was carried out using the Bio-Rad CFX Maestro 1.1 Version: 4.1.2433.1219 software (Bio-Rad, USA), as well as using the Microsoft Excel program (Microsoft, USA).

Results: During the study, we found that the activity of the studied genes of the RNA interfering system of wheat plants actually changed under the influence of the phytopathogen (Fig. 1).

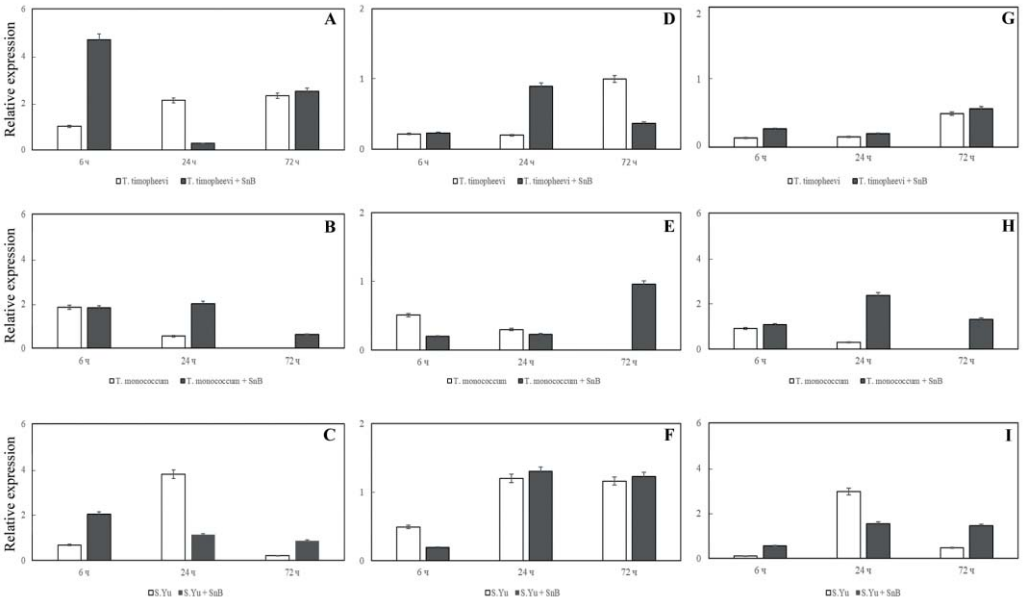


Fig. 1. Changes in transcription activity of the genes Ago1 (A, B, C), Ago2 (D, E, F) and Dcl2 (G, H, I) in *T. timopheevii* (A, D, G), *T. monococcum* (B, E, H) and *T. aestivum* variety "Salavat Yulaev" (C, F, I) under normal conditions and when infected with the pathogenic fungus *S. nodorum* 6, 24 and 72 hours after infection

It was found that in the resistant tetraploid wheat *T. timopheevii*, the activity of the Ago1 gene increased when attacked by a phytopathogen already 6 hours after infection, while the activity of Ago2 was observed only at the 24-hour fixation point, which can be explained by the sequential activation of these Ago genes and correlates with data on their relationship [2]. Similar results were observed when analyzing the studied RNAi genes in the hexaploid bread wheat variety "Salavat Yulaev", which is moderately susceptible to the Septoria pathogen. On the other hand, this variety showed a tendency to activate a gene from the Dcl family, which correlates with data previously obtained in our laboratory [3]. At the same time, the activity of the Ago and Dcl genes increased at later stages of pathogenesis in diploid *T. monococcum*.

Conclusion: Thus, our data indicate the involvement of the protein components Ago and DCL, involved in the phenomenon of RNA interference, in the protective response of wheat plants against the phytopathogen *S. nodorum*.

Funding: The study is supported by the grant No. 24-26-00266 of Russian Science Foundation.

References

1. Shein M.Y., Burkhanova G.F., Maksimov I.V. The effect of salicylic and jasmonic acids on the activity of SnAGO genes in the fungus Stagonospora nodorum Berk. in in vitro culture and during infection of wheat plants. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(8):1000-1009. doi 10.18699/VJGB-23-115
2. Shen D., Suhrkamp I., Wang Y., Liu S., Menkhaus J., Verreet J.A., Fan L., Cai D. Identification and characterization of micro RNAs in oilseed rape (*Brassica napus*) responsive to infection with the pathogenic fungus *Verticillium longisporum* using *Brassica* AA (*Brassica rap*a) and CC (*Brassica oleracea*) as reference genomes. *New Phytol*. 2014;204(3):577-594. doi 10.1111/nph.12934
3. Shein M., Maksimov I., Burkhanova G. The role of RNA interference in the formation of protective systems of wheat against the pathogen Septoria *Stagonospora nodorum* Berk. *FEBS Open Bio*. 2021;11:123

doi 10.18699/bgrs2024-06-50

Transcriptomic changes in potato after inoculation with Potato spindle tuber viroid propagated in tomato plants

Shmakov N.[1, 2]*, Afonnikov D.[1, 2], Vasiliev G.[1], Shatskaya N.[1], Egorova A.[1], Mironenko N.[3], Lashina N.[3], Khiutti A.[3], Afanasenko O.[3], Kochetov A.[1, 2, 4]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Faculty of Natural Sciences, Novosibirsk State University, Novosibirsk, Russia*

[3] *All-Russian Institute of Plant Protection, St. Petersburg, Russia*

[4] *Faculty of Agronomy, Novosibirsk State Agrarian University, Novosibirsk, Russia*

* *shmakov@bionet.nsc.ru*

Key words: plant immunity; plant transcriptomics; potato; potato spindle tuber viroid

Motivation and Aim: Potato (*Solanum tuberosum* L.) is one of the most important crops worldwide. A number of pathogens can negatively affect its yield. Among such pathogens is potato spindle tuber viroid (PSTVd), a member of Pospiviroid genus. Tomato strains of PSTVd are highly pathogenic to potato cultivars [1]. However, after three cycles of propagation in tomato, PSTVd causes only mild symptoms in potato.
Methods and Algorithms: Transcriptome of potato leaves at three time points (prior to inoculation, 14 dpi and 30 dpi) was sequenced using Illumina NextSeq 550 platform. RNA-seq analysis was performed, with reads being mapped to the *S. tuberosum* reference genome with Dart software, and differential expression assessed with EdgeR package for R language. Gene Ontology and pathway involvement analysis of differentially expressed genes was undertaken.
Results: RNA-seq analysis allowed to observe significant changes of gene expression at 14 dpi and 30 dpi. Genes with decreased expression at 30 dpi are associated with DNA replication, regulation of cell cycle and cell wall-associated and microtubule-based processes, indicating suppression of growth and cell division. Additionally, decrease in general metabolism and metabolism-related functions was observed.
Conclusion: Amount of genes with significant differential expression between 0 dpi and 30 dpi points to at a considerable transcriptomic and metabolomic reprogramming of potato plants in response to PSTVd inoculation. Potato plants demonstrated general response to abiotic stresses and initiation of dormancy and transient growth delay [2].
Funding: The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation, No. 075-15-2019-1662; this study was supported in part through computational resources of HPC facilities at collaborative center “Bioinformatics” ICG SB RAS.

References

1. Afanasenko O.S., Lashina N.M., Mironenko N.V. et al. Evaluation of responses of potato cultivars to potato spindle tuber viroid and to mixed viroid/viral infection. *Agronomy*. 2022;12(12):2916
2. Kochetov A.V., Shmakov N., Afonnikov D.A. et al. Three Cycles of Continuous Propagation of a Severe PSTVd Strain NicTr-3 in Solanum lycopersicum cv. Rutgers Resulted in Its Attenuation and Very Mild Disease Symptoms in Potato. *Agronomy*. 2023;13(3):684. doi 10.3390/agronomy13030684

doi 10.18699/bgrs2024-06-51

Genetic diversity shaped by transposon variation in chickpea

Stanin V.*, Duk M., Bankin M., Kanapin A., Samsonova A., Samsonova M.
Peter the Great Polytechnic University, St. Petersburg, Russia
** staninvladislav@mail.ru*

Key words: GWAS; chickpea; multi-locus models; agronomic traits

Motivation and Aim: Chickpea is a crucial crop for the agricultural sector and serves as a significant contributor to the economy of many developing countries. It finds use as a source of human nutrition and livestock feed alike. However, the sensitivity of chickpea to both biotic and abiotic factors has been on the rise, primarily due to the reduced genetic diversity in cultivars since the mid-twentieth century. In certain years, drought and extreme temperatures have been responsible for up to half of the global crop loss [1, 2], thereby underscoring the need for developing productive and climate-resistant chickpea genotypes. Achieving this objective would be a significant challenge for the crops industry.

Transposable elements (TEs) are mobile genomic elements that enable plants to adapt to environmental conditions. Up to 80 % of a plant cell's genome is comprised of TEs which can influence agricultural traits such as plant growth, size, color and shape of fruits or flowers, flowering time, and resistance to pests [3, 4]. Studying mobile elements in the chickpea genome will facilitate breeding and obtaining climate-resilient varieties. This study evaluates the contribution of TEs to chickpea diversity.

Methods and Algorithms: We collected 190 chickpea genotypes from the Vavilov All-Russian Institute of Plant Genetic Resources in Russia, comprising of 167 landraces collected by Vavilov in 1930s and 23 elite cultivars. We grouped the samples into seven geographic regions based on proximity (i.e., Mediterranean, Lebanon, South of Russia, Turkey, Uzbekistan, India, and Ethiopia). Short-read genome sequencing with the Illumina protocol yielded a dataset with an average coverage of 25x or about 37 Gbp per sample. The reads were processed and aligned to the chickpea reference genome assembly ASM33114v1. PopoolationTE 2 tool was used to predict mobile elements in separate- and joint-analysis modes [5].

We conducted two modes of analysis to study the transposon content of chickpea samples. In the first mode, we analyzed the transposon content of each sample separately, while in the second mode, we analyzed the transposon composition of the entire set. By obtaining lists of mobile elements and their coordinates, we were able to perform further analyses, such as identifying the distributions of TE types and the positions of their insertions in the genome with respect to gene coordinates. We also used the PrimatR statistical package to identify “hot spots” where TE insertions are densely located. Additionally, we studied TE polymorphic positions (TIPs) that are specific to chickpea populations from diverse geographical and climatic collection sites. To assess the contribution of TIPs to phenotypic variation, we conducted GWAS using both GAPIT [6] and the IIIVmrMLM program [7] and data on 12 phenotypic traits measured at two experimental stations, Astrakhan and Kuban, in 2022.

Results: Our findings indicate that the majority of transposable insertion polymorphisms belong to the Copia and Gypsy families, with 41 % and 16 %, respectively. The introns

of genes were found to have the highest frequency of polymorphic positions. 47 TE insertion hotspots were present across all chromosomes, with the lowest number of hotspots observed in chromosomes 5 and 8. Furthermore, we found that the Ethiopian population had a higher incidence of TIPs, which were predominantly found in this population. Finally, a thorough GWAS analysis identified 244 quantitative trait nucleotides (QTNs) that explained 18.1 % to 65.61 % of the variation for different phenotypic traits. The maximum amount of trait variation explained by a single QTN was 13.71 %. Of the 244 QTNs identified, 28 were mapped to genes with known function or located within the 1 kb-flanking region of these genes. These results shed new light on the genetic variation and distribution of TIPs, as well as the presence of QTNs that contribute to the phenotypic variability of certain traits.

Conclusions: Overall, our findings highlight the significant role played by TIPs in the generation of phenotypic diversity in chickpea and hold great potential for future breeding programs, offering a promising avenue for the development of improved cultivars.

Funding: This work is supported by the Russian Science Foundation grant 22-46-02004.

References

1. Kaloki P., Devasirvatham V., Tan D.K.Y. Chickpea Abiotic Stresses: Combating Drought, Heat and Cold. In: Abiotic and Biotic Stress in Plants. *IntechOpen*. 2019. doi 10.5772/intechopen.83404
2. Varshney R.K., Thudi M., Roorkiwal M. et al. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat Genet*. 2019;51(5):857-864
3. Mhiri C., Borges F., Grandbastien M.A. Specificities and Dynamics of Transposable Elements in Land Plants. *Biology (Basel)*. 2022;11(4):488
4. Oliver K.R., Greene W.K. Transposable elements: powerful facilitators of evolution. *BioEssays*. 2009;31(7):703-714
5. Kofler R., Gómez-Sánchez D., Schlötterer C. PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq. *Mol Biol Evol*. 2016;33(10):2759-2764
6. Wang J., Zhang Z. GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinformatics*. 2021;19(4):629-640. doi 10.1016/j.gpb.2021.08.005
7. Li M., Zhang Y.W., Xiang Y., Liu M.H., Zhang Y.M. IIIVmrMLM: The R and C++ tools associated with 3VmrMLM, a comprehensive GWAS method for dissecting quantitative traits. *Mol Plant*. 2022;15(8):1251-1253

doi 10.18699/bgrs2024-06-52

Rye inbred line L371 as a genetic source for primary octoploid triticale resistance to wheat leaf rust

Tsvetkova N.V.[1]*, Andreeva E.A.[1, 2], Zykin P.A.[1], Tyryshkin L.G.[3]

[1] *Saint Petersburg State University, St. Petersburg, Russia*

[2] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

[3] *Federal Research Center N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia*

* *n.tswetkowa@spbu.ru*

Key words: rye; wheat-rye hybrids; wheat leaf rust; *Puccinia triticina*

Motivation and Aim: Triticale (×Triticosecale Wittmack) is a synthetic species of the Poaceae family, artificially created more than 100 years ago by crossing of wheat and rye species, mainly *Triticum aestivum* L. and *Secale cereale* L. The greatest value of triticale is extremely wide ecological plasticity inherited from the rye parent. One of the main factors decreasing triticale yield and quality is fungal diseases. Being an allopolyploid, in theory triticale could combine diseases resistance genes of wheat and rye and therefore potentially have a greater variability resistance reaction. In practice triticale of various levels of ploidy is affected by many fungal diseases (stem and leaf rusts, root rots, powdery mildew, etc.) [1].

It has been shown that the gene pool of triticale is extremely poor in terms of effective seedling resistance to leaf rust [2] caused by the fungus *Puccinia triticina* Erikss. Susceptibility of most triticale genotypes to leaf rust could be explained by use in their creation of both (bread wheat and rye) parents susceptible to the disease.

We supposed that breeding of resistant triticale could be in some cases successful when rye resistant to leaf rust is used as one parent component. Previously, line L371 from the Peterhof's Genetic Collection [3] was described as resistant to leaf rust. This line as well as a highly susceptible line L393 were used to produce wheat-rye hybrids (ABDR amphihaploids) and amphidiploids (AABBDDRR = 2n) for their subsequent testing on resistance to leaf rust and genetic analysis to uncover the number of genes controlling leaf rust resistance. Moreover, the analysis of genetic markers linked to leaf rust resistance genes (Lr genes) was performed to understand the novelty of resistance genes from rye line L371.

Methods and Algorithms: To reveal the number of genes involved in conferring resistance to wheat leaf rust genetic analysis was performed on reciprocal hybrids between rye lines L371 and L393 as well as on wheat-rye hybrids (ABDR amphihaploids). Wheat leaf rust resistance of hybrids was evaluated with a complex population of *P. triticina*. For analysis of DNA markers of *Lr* genes PCR with specific primers [4–9], fragment analysis and sequencing of amplified fragments were done. Marker localization in reference genomes of *Aegilops tauschii* (GCF_002575655.2), *S. cereale* (GCA_902687465.1) was determined by *in silico* PCR using ipcress from the exonerate package and searching for previously sequenced sequences in these genomes using the blastn version 2.12.0. Regions of 4 million base pairs upstream of the marker and 4 million base pairs downstream the marker were used for *ab initio* automatic annotations using Augustus (https://bioinf.uni-greifswald.de/augustus/), annotated

sequences were searched in the SwissProt database (version 2024_02), from the final list of genes those related to transposable elements and genes with an e-value worse than 1×10^{-30} were removed.

Results: Evaluation of resistance to wheat leaf rust revealed that rye line L371, as well as amphihaploids and amphidiploids were highly resistant to the complex population of wheat leaf rust after inoculation of intact seedlings. L393 and Chinese Spring were highly susceptible to this inoculum.

Hybridological analysis of rye resistance to wheat leaf rust performed on two reciprocal crosses of rye lines (L371 × L393 and L393 × L371) revealed that F_1 plants from the reciprocal crosses of two rye lines were highly resistant to the rust, so the resistance is a dominant trait. Segregation in F2 plants are in correspondence with the hypotheses of three genes involved in genetic control of resistance: i) two recessive and one dominant genes or ii) one dominant and two complementary dominant genes. To distinguish these two hypotheses, a modified hybridological analysis proposed by Voylokov and Tikhenko [10] was used, and an analysis of the segregation for resistance to *P. triticina* in amphihaploid populations derived from crosses of hybrid F_1 rye plants with susceptible wheat variety was carried.

The segregation ratio for Chinese Spring × F_1 (L371 × L393) and for Chinese Spring × F_1 (L393 × L371) amphihaploid populations did not contradict to the theoretically expected ratio 7:1 for three genes (two recessive and one dominant) involved in genetic control of resistance. Additionally, these three genes control resistance in rye and in amphihaploids, hence bread wheat genome does not inhibit their expression.

Up to date a lot of rust resistance genes (*Lr*) used in wheat breeding programs were genetically mapped. Synthetic population of *P. triticina* used in the study was avirulent to some of *Lr* genes: *Lr9*, *Lr19*, *Lr24*, *Lr41* (=*Lr39*) and *Lr47*. So based on leaf rust resistance test it was not possible to decide whether three rye resistance genes are some of these *Lr* genes or not. To solve the problem specific DNA markers described for *Lr* genes [4–9] were used in PCR on DNA of resistant (L371) and susceptible (L393) rye lines. PCR analysis of DNA markers for *Lr9*, *Lr24* and *Lr47* genes did not lead to the amplification of any products on DNA of both rye lines. Based on the results it is possible to suggest that *Lr9*, *Lr24* and *Lr47* cannot be considered as candidate genes for three resistant rye genes described in the study.

PCR analysis of DNA markers for *Lr19* and *Lr41* genes led to the amplification of PCR products. Fragment analysis of PCR products for *Lr19* marker revealed no polymorphism in length of PCR products between resistant and susceptible rye lines, so *Lr19* also is not a candidate gene.

Fragment analysis of PCR products for *Lr41* marker Barc124 [8] revealed polymorphism in length of PCR products between resistant and susceptible rye lines, so it is possible to consider *Lr41* gene as a gene candidate for one of the three resistant rye genes. *Lr41* gene was introduced into bread wheat cultivars from *Ae. tauschii*. To confirm this suggestion PCR product for *Lr41* markers Barc124 [8] was sequenced and mapped on genomes of *S. cereale* and *Ae. tauschii*. A synteny in a region +/–4 Mb from the localization of Barc124 marker was observed.

Conclusion: Results of genetic analysis of the rye line L371 resistant to *P. triticina* performed on rye and wheat-rye hybrids revealed that this resistance is the dominant trait; cytoplasm has no or minor influence on the phenotypic expression of the resistance; the trait is controlled by three independent genes (two dominant and one recessive); the expression of each gene is not suppressed by bread wheat genome; analysis of DNA

markers of *Lr* genes and synteny of rye and *Aegilops* regions revealed that one of the possible gene candidate is *Lr41* and two other genes can be considered as new rye genes conferring resistance to leaf rust.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation, in accordance with agreement No. 075-15-2022-322, dated 22 April 2022.

References

1. Arseniuk E., Goral T. Triticale Biotic Stresses – Known and Novel Foes. In: Eudes F. (Ed.). Triticale. Springer, 2015:83-108
2. Hanzalová A., Bartoš P. Resistance of triticale to wheat leaf rust (Puccinia triticina). *Czech J Genet Plant Breed*. 2011;47(1):10-16
3. Andreeva E., Burlakovskiy M., Buzovkina I. et al. Genetic collections of St. Petersburg University. *Biol Commun.* 2023;68(3):199-214. doi 10.21638/spbu03.2023.308
4. Gupta S.K., Charpe A., Koul S., Prabhu K.V., Haq Q.M.R. Development and validation of molecular markers linked to an Aegilops umbellulate-derived leaf-rust-resistance gene, Lr9, for marker-assisted selection in bread wheat. *Genome.* 2005;48:823-830
5. Prins R., Groenewald J.Z., Marais G.F., Snape J.W., Koebner R.M.D. AFLP and STS tagging of Lr19, a gene conferring resistance to leaf rust in wheat. *Theor Appl Genet.* 2001;103:618-624
6. Gupta S.K., Charpe A., Koul S., Haque Q.M.R., Prabhu K.V. Development and validation of SCAR markers co-segregating with an Agropyron elongatum derived leaf rust resistance gene Lr24 in wheat. *Euphytica.* 2006;150:233-240
7. Singh S., Franks C.D., Huang L. et al. Lr41, Lr39, and a leaf rust resistance gene from *Aegilops cylindrica* may be allelic and are located on wheat chromosome 2DS. *Theor Appl Genet*. 2004;108:586-591
8. Sun X., Bai G., Carver B.F. Molecular markers for wheat leaf rust resistance gene Lr41. *Mol Breeding.* 2009;23:311-321
9. Helguera M., Khan I.A., Dubcovsky J. Development of PCR markers for the wheat leaf rust resistance gene Lr47. *Theor Appl Genet.* 2000;100:1137-1143
10. Voylokov A.V., Tikhenko N.D. Identification and localization of rye polymorphic genes specifically expressed in Triticale. In: Proceedings of the of 4th International Triticale Symposium, Alberta, Canada. Lacombe Int. Triticale Ass., 1998;290-296

doi 10.18699/bgrs2024-06-53

The transcription factor WRKY multigene family in grapevine (*Vitis vinifera*): genome-wide identification and evolution

Vodiasova E.[1]*, Sinchenko A.[1], Khvatkov P.[1], Dolgov S.[1, 2]

[1] *The Nikitsky Botanical Gardens – National Scientific Center of the RAS, Nikita, Yalta, Republic of the Crimea, Russia*

[2] *Branch of Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Puschino, Russia*

* *eavodiasova@gmail.com*

Key words: WRKY transcription factor; *Vitis vinifera*; genome-wide analyses: phylogeny; grape cultivars

Motivation and Aim: WRKY is one of the largest multigenic families of transcription factors in higher plants. These proteins are key regulators involved in the response to a variety of biotic and abiotic factors and play a pivotal role in plant growth and development [1]. Many theories have been proposed about the evolution of WRKY genes from algae to multicellular plants, but there is no consensus [2]. *Vitis vinifera* L. (grape) is one of the most widely cultivated plants for agricultural purposes. To date, several genomic analyses of WRKY genes in grape have been performed [3], but the number of genes encoding WRKY varied and seemed to depend on the version of the genomic assembly analysed. This study focuses on the identification of WRKYs by genomic analysis from different grapevine assemblies and attempts to shed light on their evolution.

Methods and Algorithms: The genome-wide analyses were performed using six grape genome assemblies. The identification of WRKY proteins was based on the detection of the WRKY domain signature using InterProScan v. 5.63-95.0. The exon-intron structure of each WRKY transcription factor-encoding gene was described according to the annotation of *V. vinifera* genomes from the NCBI and GrapeGenomics databases. A motif analysis was performed using the MEME online tool in classic mode. Phylogenetic analysis was performed using IQ-TREE v. 2.3.0. The data set included all identified *V. vinifera* WRKY protein sequences from six genome assemblies. Multiple alignment of 965 protein sequences was performed using the MAFFT v. 7.0 online service.

Results: The domain analysis identified WRKY DNA-binding domain proteins in each grape assembly. The number of proteins identified differed between the grape cultivars. A total of 115 proteins (including isoforms) were found for Cabernet Franc, 106 for Cabernet Sauvignon, 129 for Pinot Noir clone FPS123, 89 for the 12X assembly (GCA_000003745.2), 87 for the RefSeq reference assembly (GCA_030704535), 65 for the GenBank reference assembly (GCA_030704535) and 84 for the assembly from the GrapeGenomics database. The phylogenetic analysis identified 62 gene clusters with 100 bootstrap supports characterising different WRKY classes. Each WRKY class was numbered in accordance with the position of the encoding gene in the majority of the assemblies analysed. It was observed that some WRKYs were not present in all grape cultivars. The number of exons ranged from 2 to 8 and the length of VvWRKY from 151 to 746 aa. The mean genetic distances between varieties ranged from 0 to 0.113. Variability of the conserved DNA-binding heptapeptide WRKYGQK was detected. Some additional domains were identified in several grape VvWRKYs: a Zn cluster

domain (IPR018872); a Frigida-like domain (IPRO12474); the COILS structural motif and the signal peptide in the N-terminal and non-cytoplasmic domains. Four novel chimeric WRKY TFs were thus revealed. The LxLxLx repressor and LxxLL co-activator motifs were also identified in eight and twelve VvWRKY proteins, respectively. An association between the clustering of WRKY groups with the functional family of the protein, WRKY signature sequences and the presence of other domains in the protein with the formation of chimeric WRKY TFs was revealed. Based on the phylogenetic analysis performed and the probability of evolutionary events, we suggest that there were two dynamic phases of complexity and simplification in the evolution of WRKY. The two clades C and D evolved from a common ancestor as a result of the complexity of the gene structure, then divergence occurred and a characteristic motif (NTWD or Coil) emerged in each clade. During the simplification phase, loss of one of the motifs resulted in a decrease in exon number and protein length, consistent with a bimodal distribution of exon length and number. The proposed evolutionary model is consistent with the fact that macroevolutionary patterns are characterised by a periodicity of dynamics in opposite directions. Quantitatively, which is also consistent with the proposed theory, genome evolution is dominated by reduction and simplification, followed by episodes of increasing complexity.

Conclusion: In *V. vinifera*, sixty-two *WRKY* transcription factor genes have been identified. The structure of each of the *VvWRKY* genes has been studied and its chromosomal location has been determined. The genes were re-numbered by chromosome location in the reference genome of *V. vinifera* cv. PN40024, v.5. was proposed. Inter-varietal amino acid variability was revealed, reaching 5 % for some genes. In addition, chimeric *VvWRKY* genes have been found, which may have specific regulatory functions. Phylogenetic analysis shows that the evolution of *WRKY* genes went through a phase of complication (gaining introns and forming genes with two domains) and a phase of simplification (losing introns and reducing protein length). Consistent with the wide range of processes in which WRKY transcription factors are involved, the data obtained indicate a high functional flexibility of this family.

Funding: The study is supported by the Russian Science Foundation Grant (No. 23-76-10013).

References

1. Wani S.H., Anand S., Singh B., Bohra A., Joshi R. WRKY Transcription Factors and Plant Defense Responses: Latest Discoveries and Future Prospects. *Plant Cell Rep.* 2021;40(7):1071-1085. doi 10.1007/s00299-021-02691-8
2. Goyal P., Devi R., Verma B., Hussain S., Arora P., Tabassum R., Gupta S. WRKY Transcription Factors: Evolution, Regulation, and Functional Diversity in Plants. *Protoplasma.* 2023;260(2):331-348. doi 10.1007/s00709-022-01794-7
3. Wu W., Fu P., Lu J. Grapevine WRKY Transcription Factors. *Fruit Res.* 2022;2(1):1-8. doi 10.48130/frures-2022-0010

doi 10.18699/bgrs2024-06-54

Computational prediction of interactions of long non-coding RNAs and micro RNAs in maize

Yan J.T.[1]*, Pronozin A.[2, 3], Afonnikov D.[1, 2, 3]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *t.yan5@g.nsu.ru*

Key words: long non-coding RNA; microRNA; interaction; maize; prediction; computer method

Motivation and Aim: Only 1 to 2 % of the genome sequences in the human body have protein-coding functions, and the content of non-coding regions without protein-coding capabilities is as high as 98 % [1]. The remaining expressed genes are likely to be transcribed into non-coding RNA (ncRNA). However, the research on long non-coding RNA (>200 nt, lncRNAs) has just entered the development stage, and the research on plant lncRNA started even later. Therefore, the research on plant lncRNA may reveal unknown new mechanisms controlling plant growth and differentiation [2]. Micro RNA (miRNAs) are small (22–24 nt) ncRNAs that can interact with protein-coding transcripts (mRNAs) providing their subsequent degradation and regulating their functioning. Interactions between miRNAs and mRNAs has been studied in detail [3]. However, the interaction between miRNA and lncRNA has rarely been studied in plants. So identification of potential MIRNA-lncRNA interactions in various plant species remains a challenging task. Furthermore, among the identified miRNA-lncRNA interactions, only a few have been analyzed for biological significance experimentally. The detailed functional mechanisms of these interactions remain unclear. It will also be interesting to understand whether these miRNA-lncRNA interactions, related functions, and mechanisms are conserved between different plant species. At present, lncRNA and miRNA discovered in animals and plants are difficult to verify through large-scale biological experiments. Therefore, there is an urgent need to use computational methods to reveal the characteristics of lncRNAs and miRNAs to guide those expensive and laborious laboratory experiments [1].

Methods and Algorithms: Most research on the mutual regulation mechanism between miRNA and lncRNA focuses on animal and human cancers, with relatively few studies on plants, and even less on corn. In order to deeply explore the interaction between corn miRNA and lncRNA, this article uses the PmliPred method [4]. This method is based on hybrid models and fuzzy decision-making, combines deep learning and shallow machine learning, and uses original sequences and artificial extraction of characteristic plant miRNA-lncRNA to achieve classification prediction of miRNA-lncRNA interaction relationships. Through the PmliPred method, the interaction information between miRNA and lncRNA can be used to predict potential relationships in the maize gene regulatory network. In order to achieve this goal, we first collect data, then preprocess the data and write a series of python programs to extract features from the data. PmliPred is then run for predictive analysis based on their sequences and features. As an effective tool, PmliPred helps us deeply explore the complex mechanisms of gene regulation in

maize, better predict the interaction between miRNA and lncRNA in maize, and provide an important reference for further functional research and breeding.

Results: Using PmliPred software, we predicted putative interaction between maize lncRNAs [5] and miRNAs. This allowed to reconstruct graph of the miRNA-lncRNA interacting pairs. Analysis of co-expression of the lncRNAs involved in the various interactions with miRNAs was also performed.

Conclusion: Our research provide new information about miRNA-lncRNA interaction in maize.

Funding: This work was funded by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation No. 075-15-2019-1662.

References

1. Sheng N., Huang L., Gao L., Cao Y., Xie X., Wang Y. A Survey of Computational Methods and Databases for lncRNA-MiRNA Interaction Prediction. *IEEE/ACM Trans Comput Biol Bioinform*. 2023;20(5):2810-2826. doi 10.1109/TCBB.2023.3264254
2. Huang X.Q., Li D.D., Wu J. Long non-coding RNAs in plants. *Yi Chuan*. 2015;37(4):344-359. doi 10.16288/j.yczz.14-432 (in Chinese)
3. Yoon J.H., Abdelmohsen K., Gorospe M. Functional interactions among microRNAs and long noncoding RNAs. *Semin Cell Dev Biol*. 2014;34:9-14. doi 10.1016/j.semcdb.2014.05.015
4. Kang Q., Meng J., Cui J., Luan Y., Chen M. PmliPred: a method based on hybrid model and fuzzy decision for plant miRNA-lncRNA interaction prediction. *Bioinformatics*. 2020;36(10):2986-2992. doi 10.1093/bioinformatics/btaa074
5. Pronozin A.Y., Afonnikov D.A. ICAnnoLncRNA: A Snakemake Pipeline for a Long Non-Coding-RNA Search and Annotation in Transcriptomic Sequences. *Genes (Basel)*. 2023;14(7):1331. doi 10.3390/genes14071331

doi 10.18699/bgrs2024-06-55

Cost-effective estimation of optimal number of reads in GBS sequencing

Zamalutdinov A.*, Boldyrev S., Ben C., Gentzbittel L.
Skolkovo Institute of Science and Technology, Moscow, Russia
** Alexey.Zamalutdinov@skoltech.ru*

Key words: GBS; read number; soybean

Motivation and Aim: Genotype-by-sequencing (GBS) is a cost-effective approach for large-scale genotyping that has been widely employed for different species, particularly those with large genomes. In order to reduce genome complexity, GBS uses digestion of the genome by one or more restriction enzymes [1]. They are different in their properties and affect the size, number, and genome localization of the fragments. That is why, the choices of the restriction enzymes for library preparation are a crucial step in GBS. Several protocols and different enzyme combinations were evaluated in order to reduce costs and increase accuracy [2, 3]. However, quality and genotyping cost also depend on number of reads used [4]. Here, we addressed this question and suggested cost-effective empirical approach to define range for optimal number of reads.
Methods and Algorithms: 12 soybean accessions were genotyped using GBS. HindIII-NlaIII enzyme combination was used for library preparation. The resulting library was sequenced using the Illumina HiSeq4000 in a 2*150bp paired-end mode with 111 million paired-end reads. In order to evaluate the dependency between number of reads and number of SNPs, random sampling of reads was performed in triplicate, from 1 million reads per library up to 111 million. Demultiplication and SNP calling were performed using stacks with default settings [5]. Trimming and filtering were conducted using the bbduk tool [6] with following parameters: "k=31 ref=artifacts,phix,adapters,lambda,pjet,mtst,kapa ordered cardinality qtrim=rl trimq=20 maq=25 minlen=50 tbo mink=11 ktrim=r minlen=50". The mapping was performed using BWA-MEM [7]. The vcf files were subjected to filtering using Bcftools 1.18 [8], with genotypes with depth < 3 set to missing. Subsequently, biallelic single-nucleotide polymorphisms (SNPs) with an average depth per genotype of greater than five, a minor allele frequency (MAF) of 0.05, and a fraction of missing data of less than 0.4 were used for further analysis. In addition, we estimated the fraction of genome regions covered by average number of reads per accession in each library. Fraction of covered regions was calculated using Bedtools 2.30.0 [9].
Results: Based on real sequencing data obtained through the use of HindIII-NlaIII combination, libraries of different sizes were obtained by random read selection. It is expected that more reads are used, more SNPs are obtained. However, this dependency is not linear and three regions on the curve can be identified (Fig. 1A). Low number of reads will output extremely low number of SNPs. In the middle linear dependency between reads and SNPs count is observed. High number of reads will generate similar number of SNPs. So, here it is clear how many SNPs will be called using this number of reads. However, it requires preparation of small library of several accessions with high sequencing depth.

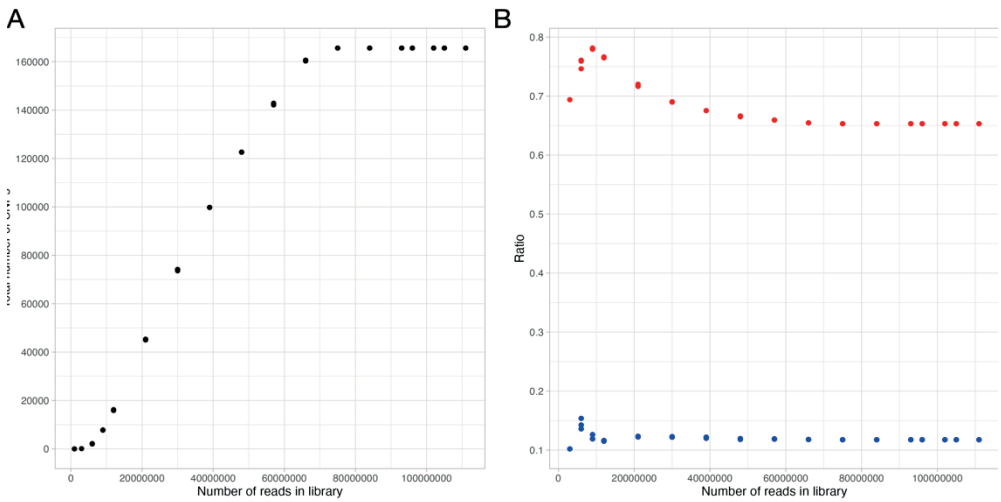


Fig. 1. A – Dependence of the total number of SNPs on the number of reads for library sequencing. Three regions of interest can be identified on the curve. From 1 to 6 million reads there is no increase in the number of SNPs. From 9 to 66 million – linear dependence, more reads are used, more SNPs are obtained. From 75 to 111 there is no increase in the number of SNPs, only an increase in depth and accuracy. B – Dependency of the fraction of SNPs in repetitive regions and in genes on the number of reads for library sequencing. The proportion of SNPs in repeats is depicted in red, while the proportion of SNPs in genes is shown in blue. The ratio of SNPs in genes is stable from 9 million reads per library. In contrast, the proportion of SNPs in repeats decreases slightly as the number of reads used increases

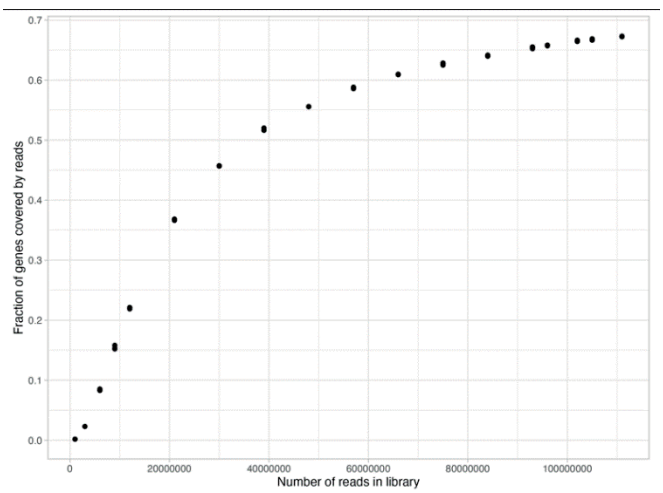


Fig. 2. Dependency of fraction of genes covered on number of reads for library sequencing. First additions of read number rapidly increase the fraction of covered genes. However, the maximum possible number of covered genes is predefined by library design, or enzyme choice. So, at high numbers of used reads fraction of covered genes increases slowly

We also examined the localization of SNPs in genome and find that fraction of SNPs in genes is quite stable in different read numbers. The fraction of SNPs in repeats is decrease slowly and reach plateau at 75 million reads, together with total number of reads (Fig. 1B).

Nevertheless, SNP calling requires relatively high coverage for each sample and depends on collection diversity. We evaluated which information can be used to estimate required number of reads without sequencing of libraries, but using only one sample based on genome features. Reads were sampled according to average number of reads per accession in examined libraries and the fraction of genes covered by different numbers of reads was estimated. Figure 2 demonstrates that it is again non-linear dependency on read number. At small number of reads rapid growth of covered genes can be observed. It corresponds to the first region form Figure 1A. However, saturation of covered genes is rather smooth and no sharp bend can be observed. Huge number of reads will not significantly increase the fraction of covered genes as maximum number of possibly covered genes is predefined by enzyme choice.

Conclusion: GBS genotyping is well-established approach and widely used for many plant species. However, the results depend on many factors, such as enzyme choice, library preparation, and read number. Here, we demonstrated that read number affects final number of SNPs as well as their localization in soybean using a real dataset with the HindIII-NlaIII enzyme combination. However, estimating the optimal number of reads by genotyping the test GBS library is either costly when high coverage is utilized or inaccurate when low numbers of reads are used. Alternative approach involving high coverage genotyping of the one sample without SNP calling was suggested. Such indirect approach allows to define the range of read numbers where costs will be adequate to output. We consider that this approach provides easy way to estimate adequate read number per library and can be extended to other species and enzyme combinations.

References

1. Elshire R.J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011;6(5):e19379. doi 10.1371/journal.pone.0019379
2. de Ronne M., Légaré G., Belzile F., Boyle B., Torkamaneh D. 3D-GBS: a universal genotyping-by-sequencing approach for genomic selection and other high-throughput low-cost applications in species with small to medium-sized genomes. *Plant Methods*. 2023;19(1):13. doi 10.1186/s13007-023-00990-7
3. Hamblin M.T., Rabbi I.Y. The effects of restriction-enzyme choice on properties of genotyping-by-sequencing libraries: A study in Cassava (*Manihot esculenta*). *Crop Sci*. 2014;54(6):2603-2608. doi 10.2135/cropsci2014.02.0160
4. Beissinger T.M. et al. Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics*. 2013;193(4):1073-1081. doi 10.1534/genetics.112.147710
5. Catchen J., Hohenlohe P.A., Bassham S., Amores A., Cresko W.A. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 2013;22(11):3124-3140. doi 10.1111/mec.12354
6. Bushnell B., Rood J., Singer E. BBMerge – Accurate paired shotgun read merging via overlap. *PLoS One*. 2017;12(10):e0185056. doi 10.1371/journal.pone.0185056
7. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013. doi 10.48550/arXiv.1303.3997
8. Danecek P. et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008. doi 10.1093/gigascience/giab008
9. Quinlan A.R., Hall I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi 10.1093/bioinformatics/btq033

doi 10.18699/bgrs2024-06-56

Identifying trichomes on digital images of soy leaves

Zhang X.Y.[1]*, Menkov M.T.[2, 3], Rozanova I.V.[3], Afonnikov D.A.[1, 4]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Sirius University of Science and Technology, Sirius Federal Territory, Krasnodar region, Russia*

[3] *N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia*

[4] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

** s.chzhan7@g.nsu.ru*

Key words: soy; leaf hairiness; trichomes; image analysis; index-based segmentation; plant/background segmentation; thresholding segmentation

Motivation and Aim: Trichomes are extensions of the aforementioned epidermal cells in plants [1]. Soybean leaf trichome plays an important biological role in resisting pests and adapting to the environment and exhibits a wide range of phenotypic variation. Currently, the typical method used to collect trichome data is qualitative visual counting using lenses, which is labor- and time-intensive [2], another method is the use of scanning electron microscope(SEM), which is costly, time-consuming and not conducive to high-throughput analysis [3]. As more and more plant and trait parameters need to be measured quickly and accurately, various types of computer vision algorithms, image processing and machine learning classification methods are increasingly used in plant phenomics data analysis [4–6]. In this work, we suggest method for an automated identification of trichomes on the surface of soybean leaves through machine vision segmentation algorithms.

Methods and Algorithms: With the use of computer technology for analysis of microscopic images of transverse fold of leaf blades has been proposed for quantitative evaluation of leaf trichomes [7]. Typical image contains green leaf fold with darker trichomes located outside on the black background. Additionally, part of image contain image of ruler and labelling data. We developed an algorithm, which enables automatic segmentation of image into four parts: background, leaf, trichomes and ruler/label.

Three algorithms was used for image segmentation: (1) using pixel intensity in the RGB color space; (2) using pixel ExG, ExR, ExGR indices [8]; (3) using thresholds in H (HSV color space) and a (Lab color space). The performance of pixel classification was estimated using TP, TN, FP, FN, F1 measures. The algorithms implemented using OpenCV library.

Results: With the help of pre-segmentation data, segmentation of pixel points of different classes is performed for automatic calculation of threshold value. After completing the automatic segmentation algorithm, the results are evaluated using the confusion matrix for the multicategorization task.

Conclusion: By analyzing the number, length and other characteristic of trichomes, we will explore the association between these microstructures and the function and growth status of soybean leaves, and further reveal the important role of trichomes in plant adaptation and growth and development. By combining machine vision counting and plant phenomics research, we expect to provide important data support for geneticists to study soybean growth and development.

Funding: The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, in agreement with the Ministry of Education and Science of the Russian Federation, No. 075-15-2019-1662.

References

1. Johnson H.B. Plant pubescence: An ecological perspective. *Bot Rev*. 1975;41:233-258. doi 10.1007/BF02860838
2. Pshenichnikova T.A., Lapochkina I.F., Shchukina L.V. The inheritance of morphological and biochemical traits introgressed into common wheat (*Triticum aestivum* L.) from *Aegilops speltoides* Tausch. *Genet Resour Crop Evol.* 2007;54:287-293
3. Luo D., Oppenheimer D.G. Genetic control of trichome branch number in Arabidopsis: the roles of the FURCA loci. *Development*. 1999;126:5547-5557
4. Tsaftaris S.A., Minervini M., Scharr H. Machine learning for plant phenotyping needs image processing. *Trends Plant Sci.* 2016;21(12):989-991. doi 10.1016/j.tplants.2016.10.002
5. Pound M.P., Atkinson J.A., Townsend A.J. et al. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *Gigascience.* 2017;6(10):1-10.
6. King A. Technology:the future of agriculture. *Nature*. 2017;544(7651):S21-S23. doi 10.1038/544S21a
7. Genaev M.A., Doroshkov A.V., Pshenichnikova T.A. et al. Extraction of quantitative characteristics describing wheat leaf pubescence with a novel image-processing technique. *Planta*. 2012;236:1943-1954. doi 10.1007/s00425-012-1751-6
8. Riehle D., Reiser D., Griepentrog H.W. Robust index-based semantic plant/background segmentation for RGB-images. *Comput Electron Agric*. 2020;169:105201. doi 10.1016/j.compag.2019.105201

7 Симпозиум «Генетика/ геномика, биоинформатика и системная биология животных»

Symposium “Genetics/genomics, bioinformatics and systems biology of animals”

doi 10.18699/bgrs2024-7.1-01

Поиск и анализ последовательностей, кодирующих фенолоксидазы эндемичного вида байкальских амфипод *Eulimnogammarus verrucosus*

Арбузова Г.*, Дроздова П., Золотовская Е., Помазкин В., Тимофеев М.
НИИ биологии, Иркутский государственный университет, Иркутск, Россия
** halina.arbuzova@gmail.com*

Ключевые слова: озеро Байкал; амфиподы; фенолоксидаза

Мотивация и цель: Озеро Байкал – уникальный пресноводный водоем с высокой степенью эндемизма и продолжающимся процессом видообразования. Типичным эндемичным обитателем литорали озера является вид байкальских амфипод *Eulimnogammarus verrucosus*, занимающий важное место в его трофической структуре. Он выступает удобным модельным объектом для изучения экологии и особенностей метаболизма байкальской фауны.
Одним из способов иммунного ответа ракообразных является синтез белков-фенолоксидаз. Данные белки отвечают за меланизацию поврежденных тканей и инкапсуляцию патогенов. Фенолоксидазы ракообразных широко исследуются, но в настоящее время в литературе нет информации о белках этого семейства у амфипод. Исследование направлено на поиск генов, кодирующих фенолоксидазы *E. verrucosus*. Изучение этих белков позволит расширить фундаментальную базу знаний о метаболизме и иммунитете байкальских амфипод. Кроме того, холодоадаптированные фенолоксидазы могут быть перспективны в фармацевтической промышленности для синтеза коммерческого меланина в качестве биологически активной добавки к пище, УФ-фильтра в косметических средствах, а также для производства биосовместимых полупроводников.
Методы и алгоритмы: Для исследования генов, кодирующих фенолоксидазы *E. verrucosus*, был проведен анализ литературных источников с целью поиска консенсусных последовательностей для нахождения сходств в транскриптоме данного вида. За основу исследования были взяты известные гены, кодирующие фенолоксидазы видов *Macrobrachium nipponense* (AFA26602) и *Homarus gammarus* (CAE46724) [1]. Поиск в сборке транскриптома *E. verrucosus* осуществлен при помощи программы GenExploT [3]. Анализ последовательностей производился с использованием программы SnapGene Viewer (https://www.snapgene.com/snapgene-viewer). Для дальнейшего исследования и подбора праймеров выбирали области генов с открытой рамкой считывания и длиной аминокислотных последовательностей, соответствующих размерам известных фенолоксидаз. Для выбора плазмидного вектора и подходящих сайтов рестрикции с последовательностями проводили ПЦР *in silico* в программе UGENE [2]. Подбор праймеров осуществляли в программе SnapGene Viewer. ПЦР проводили при помощи коммерческих реакционных смесей 5x Screen Mix (Evrogen), 5x q-PCR Mix-HS (SYBR), 2.5x Mix («Синтол»), 2x Screen Mix Hs-Taq Color («БиоЛабМикс»).

Результаты: Для исследования были выбраны две наиболее подходящие последовательности нуклеотидов *E. verrucosus* – GJDV01086918.1 (2437 п. н.) (рис. 1) и GJDV01086920.1 (4404 п. н.) (рис. 2), именуемые далее EvePO1 и EvePO2 соответственно. В обеих последовательностях имеется открытая рамка считывания длиной 2187 п. н. (728 а. к.).

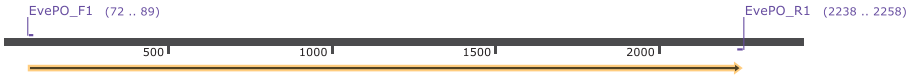


Рис. 1. Последовательность EvePO1 *E. verrucosus* (SnapGene Viewer)

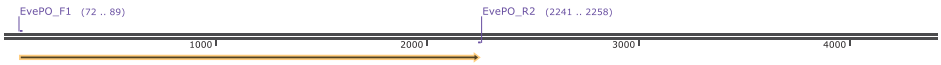


Рис. 2. Последовательность EvePO2 *E. verrucosus* (SnapGene Viewer)

Для конструирования плазмидного вектора были выбраны плазмида pPROEX и сайты рестрикции BamHI и PstI, не имеющие повторов в анализируемых последовательностях (рис. 3). Для дальнейшей работы были заказаны коммерческие праймеры, представленные ниже.

Последовательность	Прямой праймер (5’-3’)	Обратный праймер (5’-3’)
EvePO1	TCCCGGATCCATGGCGTC TACCGAGCAA	GGTCCTGCAGCTACATATCTGTG GCAGGATT
EvePO2	TCCCGGATCCATGGCGTC TACCGAGCAA	ATCCCTGCAGCTACGCATCTGTGC CTGG

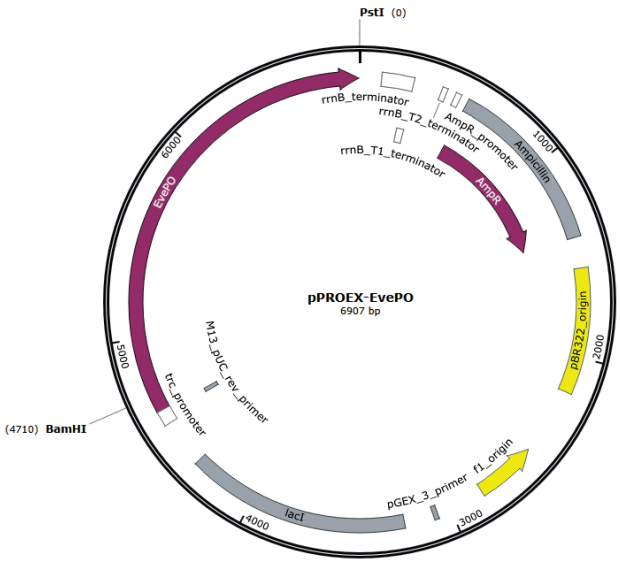


Рис. 3. Плазмидный вектор, содержащий участок гена для вставки анализируемых последовательностей, кодирующих фенолоксидазы *E. verrucosus* (SnapGene Viewer)

С кДНК *E. verrucosus* проводили амплификацию последовательностей EvePO1 и EvePO2 в градиенте температур 50–60 °C, используя праймеры EvePO_F1, EvePO_R1, EvePO_R2. На данный момент ПЦР не дала результатов (рис. 4).

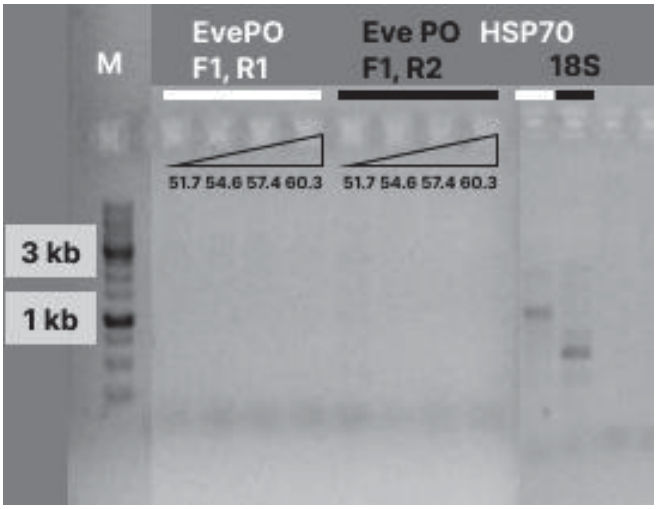


Рис. 4. Результаты ПЦР, проведенной со сконструированными праймерами к потенциальным последовательностям фенолоксидаз *E. verrucosus.* M – маркер; F1 – прямой праймер; R1, R2 – обратные праймеры

Выводы: В ходе исследования генов фенолоксидаз эндемичных байкальских амфипод *E. verrucosus* был осуществлен анализ литературных источников и поиск последовательностей, сходных с известными, кодирующими фенолоксидазы других ракообразных. Подобраны праймеры для ПЦР и сконструирован плазмидный вектор *in silico*, необходимый для дальнейшей работы. В транскриптоме *E. verrucosus* найдены как минимум две последовательности, подходящие на роль генов фенолоксидаз, для которых продолжается биоинформатический и лабораторный анализ.

Финансирование: Исследование проведено при поддержке гранта Министерства науки и высшего образования России № FZZE-2024-0008.

Search and analysis of sequences encoding phenoloxidases of the endemic Baikal amphipod species *Eulimnogammarus verrucosus*

Arbuzova G.*, Drozdova P., Zolotovskaya E., Pomazkin V., Timofeyev M.
Institute of Biology, Irkutsk State University, Irkutsk, Russia
** halina.arbuzova@gmail.com*

Key words: Lake Baikal; amphipoda; phenoloxidase

Motivation and Aim: Lake Baikal is a unique freshwater body with a high degree of endemism and an ongoing process of speciation. Typical endemic inhabitant of the littoral is the Baikal amphipod species *Eulimnogammarus verrucosus*, which plays an important role in the trophic structure of the lake. This species is a convenient model

object for studying the ecology and metabolic peculiarities of the Baikal fauna. One of the mechanisms of immune response in crustaceans is the synthesis of phenoloxidase proteins. These proteins are responsible for melanization of damaged tissues and encapsulation of pathogens. Crustacean phenoloxidase family proteins have been widely investigated, but currently there is no information on amphipod phenoloxidases in the literature. This study aims to find genes encoding phenoloxidases in the endemic Baikal amphipod species *E. verrucosus*. The study of these proteins will expand the fundamental knowledge on metabolism and immunity of Baikal amphipods. In addition, cold-adapted phenoloxidases may be used in the pharmaceutical industry for the synthesis of commercial melanin as a biologically active food additive, UV filter in cosmetics, as well as for the production of biocompatible semiconductors.

Methods and Algorithms: To investigate the genes encoding phenoloxidases of *E. verrucosus*, the available literature was analyzed to search for consensus sequences to find similarities in the transcriptome assembly of this species. The known sequences encoding phenoloxidases of *Macrobrachium nipponense* (AFA26602) and *Homarus gammarus* (CAE46724) species were taken as the query [1]. The search in the *E. verrucosus* transcriptome was carried out using the GenExploT program [3]. Sequence analysis was performed using the SnapGene Viewer program (https://www.snapgene.com/snapgene-viewer). Sequences with an open reading frame and a suitable length of the encoded amino acid sequence were selected for further study and primer selection. The selected sequences were further subjected to *in silico* PCR procedure in UGENE [2] program for selection of plasmid vector and suitable restriction sites. The selection of primers was performed in the SnapGene Viewer program. PCR of the desired sequences was performed using commercial PCR mixes 5x Screen Mix (Evrogen), 5x q-PCR Mix-HS (SYBR), 2.5x Mix ("Syntol"), 2x Screen Mix Hs-Taq Color ("BioLabMix").

Results: The two best-fitting nucleotide sequences of *E. verrucosus*, GJDV01086918.1 (2437 bp) (Fig. 1) and GJDV01086920.1 (4404 bp) (Fig. 2), hereafter referred to as EvePO1 and EvePO2 respectively, were selected for further analysis. Both sequences contain a 2187 bp (728 aa) long open reading frame.

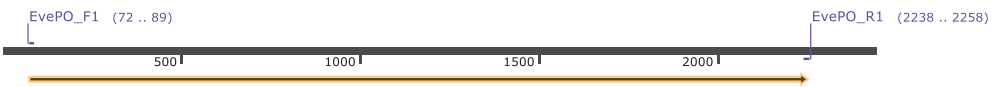


Fig. 1. EvePO1 sequence of *E. verrucosus* (SnapGene Viewer)

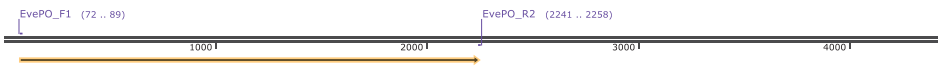


Fig. 2. EvePO2 sequence of *E. verrucosus* (SnapGene Viewer)

For plasmid construction we chose the plasmid pPROEX and the restriction sites BamHI and PstI, which are absent from the analyzed sequences (Fig. 3). Commercial primers were ordered for further work with the sequences:

Sequence	Forward primer (5’-3’)	Reverse primer (5’-3’)
EvePO1	TCCCGGATCCATGGCGTCTACC GAGCAA	GGTCCTGCAGCTACATATCTGTGGCA GGATT
EvePO2	TCCCGGATCCATGGCGTCTACC GAGCAA	ATCCCTGCAGCTACGCATCTGTGCCT GG

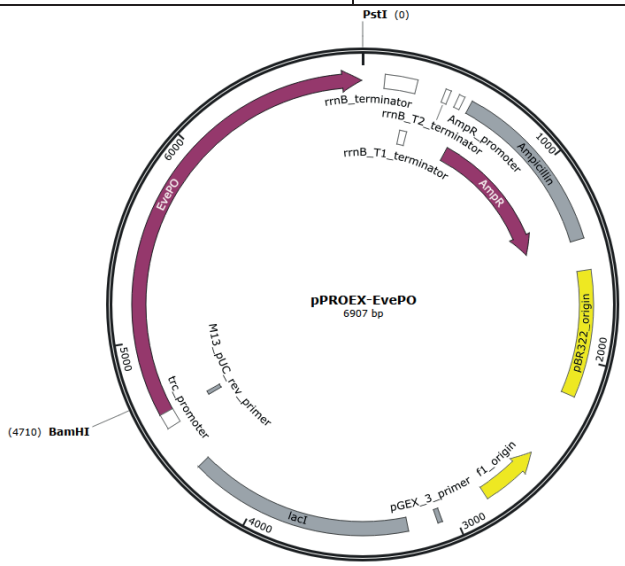


Fig. 3. Plasmid vector designed for insertion of analyzed sequences encoding phenoloxidases of *E. verrucosus* (SnapGene Viewer)

Amplification in a temperature gradient of 50-60°C was performed with the investigated sequences using cDNA of *E. verrucosus*. At this point, PCR was inconclusive (Fig. 4).

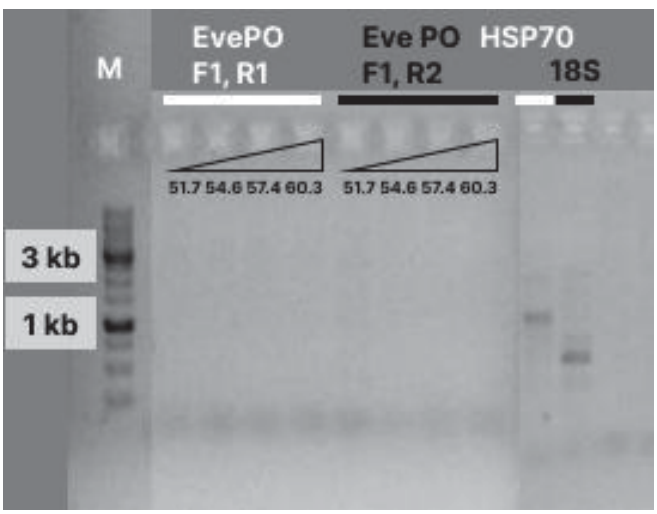


Fig. 4. PCR results of candidate sequences encoding phenoloxidases of *E. verrucosus*. Denotes: M – marker; F1 – forward primer; R1, R2 – reverse primers

Conclusion: During the study of sequences encoding phenoloxidases of the endemic Baikal amphipod *E. verrucosus*, we analyzed literature sources and searched for sequences similar to those known to encode phenoloxidases of other crustaceans. PCR primers were selected and an *in silico* plasmid vector was constructed for further work with the sequences. At least two sequences were found in the transcriptome of *E. verrucosus* that are suitable for the role of phenoloxidase encoding sequences and further bioinformatic and laboratory analysis is in progress.

Funding: This study was supported by grant No. FZZE-2024-0008 from the Russian Ministry of Science and Higher Education.

Список литературы/References

1. Arockiaraj J. et al. A novel prophenoloxidase, hemocyanin encoded copper containing active enzyme from prawn: Gene characterization. *Gene*. 2013;524(2):139-151. doi 10.1016/j.gene.2013.04.044
2. Okonechnikov K., Golosova O., Fursov M.; UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28(8):1166-1167. doi 10.1093/bioinformatics/bts091
3. Мутин А.Д. Веб-приложение для количественного анализа экспрессии и поиска гомологов транскриптов и белков у байкальских эндемичных амфипод. В: Эколого-биологические и географические исследования в решении региональных проблем. Улан-Удэ: Изд-во Бурятского университета, 2022;103-107. doi 10.18101/978-5-9793-1803-5-2022-1-186

doi 10.18699/bgrs2024-7.1-02

Размер центромер не влияет на супрессию мейотической рекомбинации вблизи центромер

Гришко Е.О.*, Котельников А.А., Слободчикова А.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
** grishko@bionet.nsc.ru*

Ключевые слова: центромера; рекомбинация; метаполицентромера; центромерный эффект

Мотивация и цель: Центромерный эффект, наряду с интерференцией между хиазмами и наличием облигатной хиазмы, является одним из универсальных свойств рекомбинации [1, 2]. Он представляет собой супрессию мейотической рекомбинации в центромерных и перицентромерных областях хромосом. На данный момент неизвестно, от чего зависит размер области подавления рекомбинации вблизи центромер. В этой работе мы проверяем гипотезу о влиянии размера центромер на силу центромерного эффекта.
Методы и алгоритмы: Мы оценили силу центромерного эффекта у четырех видов вьюрков, два из которых (обыкновенный щегол (*Carduelis carduelis*) и чиж (*Spinus spinus*)) несут на своих хромосомах региональные центромеры, а два других (обыкновенная коноплянка (*Linaria cannabina*) и обыкновенный снегирь (*Pyrrula pyrrula*)) несут метаполицентромеры – протяженные центромеры, состоящие из нескольких последовательно расположенных доменов центромерного гетерохроматина, функционирующих как одна центромера. У коноплянки метаполицентромера занимает до 5 % хромосомы, а у снегиря – до 10 %.
Исследование проводилось на препаратах пахитенных сперматоцитов, окрашенных антителами к комплексу центромерных белков, MLH1 – маркеру поздней рекомбинации и SYCP3 – белку латерального элемента синаптонемного комплекса. Мы оценили силу центромерного эффекта как среднее расстояние от края центромеры до ближайшего рекомбинационного узелка, маркированного иммуносигналом MLH1, на плечах бивалента хромосомы 1. Мы выбрали этот бивалент, поскольку смогли однозначно идентифицировать его по относительной длине и центромерному индексу у всех четырех видов.
Результаты: Расстояние от центромеры до ближайшего сигнала MLH1 уменьшается с увеличением количества сигналов MLH1 на плече. Единичные сигналы обычно располагаются в перителомерных участках вблизи концов плеч. Чем больше сигналов возникает на плече, тем ближе к центромере располагаются проксимальные сигналы. Во многих клетках с множественными сигналами MLH1 проксимальный из них располагался на границе центромерной области или очень близко к ней. Виды с региональными и метаполицентромерами не различались по среднему расстоянию от центромеры до ближайшего сигнала MLH1.
Выводы: У рассмотренных видов большее количество хиазм на плечо приводит к тому, что проксимальные хиазмы образуются ближе к центромере. Мы предполагаем, что это можно объяснить заметным сдвигом в сторону теломерного распределения рекомбинационных событий. По-видимому, у исследованных птиц рекомбинация систематически инициируется в субтеломерных областях, а интерференция смещает последующие сайты рекомбинации от теломер к

центромерам. Таким образом, дефицит сайтов рекомбинации вокруг центромеры не требует особого супрессивного действия со стороны самой центромеры, хотя и не исключает его.
Финансирование: Исследование поддержано Российским научным фондом, грант № 23-24-00304.

Centromere size does not affect the suppression of meiotic recombination near centromeres

Grishko E.O.*, Kotelnikov A.A., Slobodchikova A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** grishko@bionet.nsc.ru*

Key words: centromere; recombination; metapolycentromere; centromere effect

Motivation and Aim: The centromere effect, along with interference between chiasmata and the presence of obligate chiasma, is one of the universal properties of recombination [1, 2]. It represents the suppression of meiotic recombination in the centromeric and pericentromeric regions of chromosomes. At the moment, it is unknown what determines the size of the region of recombination suppression near centromeres. In this work, we test the hypothesis that centromere size influences the strength of the centromere effect.
Methods and Algorithms: We assessed the strength of the centromere effect in four species of finches, two of which (the goldfinch (*Carduelis carduelis*) and the siskin (*Spinus spinus*)) carry regional centromeres on their chromosomes, and the other two (the common linnet (*Linaria cannabina*) and the common bullfinch (*Pyrrula pyrrula*)) carry metapolycentromeres – extended centromeres consisting of several sequentially located domains of centromeric heterochromatin, functioning as one centromere. In the linnet, the metapolycentromere occupies up to 5 % of the chromosome, and in the bullfinch – up to 10 %.
The study was carried out on preparations of pachytene spermatocytes stained with antibodies to the complex of centromeric proteins, MLH1 – a marker of late recombination and SYCP3 – a protein of the lateral element of the synaptic complex. We estimated the strength of the centromere effect as the average distance from the centromere edge to the nearest recombination node marked by the MLH1 immunosignal on the bivalent arms of chromosome 1. We selected this bivalent because we could uniquely identify it by relative length and centromere index in all four species.
Results: The distance from the centromere to the nearest MLH1 signal decreases with increasing number of MLH1 signals on the arm. Single signals are usually located in peritelomeric regions near the ends of the arms. The more signals appear on the arm, the closer to the centromere the proximal signals are located. In many cells with multiple MLH1 signals, the proximal one was located at or very close to the border of the centromeric region. Species with regional and metapolycentromeres did not differ in the average distance from the centromere to the nearest MLH1 signal.
Conclusion: In the species considered, a larger number of chiasmata per arm leads to the fact that proximal chiasmata are formed closer to the centromere. We hypothesize that this can be explained by a marked shift toward the telomeric distribution of recombination events. Apparently, in the birds studied, the initiation of recombination is systematically shifted towards subtelomeric regions, and interference shifts subsequent

recombination sites from telomeres to centromeres. Thus, the deficiency of recombination sites around the centromere does not require a special suppressive effect on the part of the centromere itself, although it does not exclude it.

Funding: The study is supported by the Russian Science Foundation, grant No. 23-24-00304.

Список литературы/References

1. Beadle G.W. A Possible influence of the spindle fibre on crossing-over in Drosophila. *Proc Natl Acad Sci USA*. 1932;18:160-165
2. Mather K. Crossing over and heterochromatin in the X chromosome of Drosophila melanogaster. *Genetics*. 1939;24:413

doi 10.18699/bgrs2024-7.1-03

Использование бесшовной системы CRISPR/Cas9 для получения мух *Drosophila melanogaster* с точечными мутациями

Жукова М.[1]*, Яковлев К.[2], Шидловский Ю.[1, 3]

[1] *Институт биологии гена РАН, Москва, Россия*

[2] *Национальный научный центр морской биологии им. А.В. Жирмунского ДВО РАН, Владивосток, Россия*

[3] *Первый Московский государственный медицинский университет имени И.М. Сеченова, Москва, Россия*

* *zhukovamv@gmail.com*

Ключевые слова: сайт-направленный мутагенез; CRISPR/Cas9; дрозофила; генная инженерия

Мотивация и цель: Нашей задачей было получение линии *Drosophila melanogaster* с одной точечной заменой в гене *orb*, приводящей к появлению стоп-кодона и преждевременной остановке синтеза белка. Такая мутация необходима для последующего изучения функциональной роли неконсервативной С-концевой последовательности белка Orb.

Методы и алгоритмы: Бесшовная система редактирования генома CRISPR/Cas9 основана на бесследном вырезании селективных маркеров из сайта интеграции с помощью piggyBac (PB) транспозазы. PB транспозоны встраиваются по последовательностям TTAA в геноме. TTAA сайты дуплицируются при встройке транспозона и остаются в одной копии после вырезания транспозона при участии PB транспозазы. В коммерческом донорном векторе pHD-ScarlessDsRed для бесшовной системы редактирования селективный маркер *dsRed* фланкирован концами PB транспозона и TTAA сайтами, что обеспечивает его вырезание с помощью PB транспозазы. При подборе сайта разрезания генома Cas9-эндонуклеазой важно, чтобы сайт находился в непосредственной близости от TTAA сайта встройки, в этом случае последовательности донорного вектора и геномной последовательности будут отличаться и Cas9-эндонуклеаза будет разрезать только геномную ДНК, а не донорный вектор. Мы подобрали сайт для разреза ДНК на расстоянии 252 нуклеотида от проводимой точечной замены. В работе использовались стандартные методы, применяемые в генной инженерии, такие как ПЦР и молекулярное клонирование. Для получения точечной мутации мы вводили соответствующую замену в одно из двух плеч, окружающих селективный маркер *dsRed* в донорном векторе. Точечную замену вводили в два перекрывающихся комплиментарных праймера, амплифицировали с этими праймерами два продукта ПЦР. Получившиеся продукты смешивали и проводили сшивающую ПЦР, при которой используются праймеры, комплиментарные 5’ и 3’ концам суммарного фрагмента. В ранние эмбрионы линии мух, экспрессирующей эндонуклеазу Cas9 (линия #55821 из Bloomington Drosophila Stock Center, Блумингтон, США), инъецировали водный раствор ДНК, содержащий донорный вектор и плазмиду, экспрессирующую sgРНК. С развившимися из эмбрионов мухами проводили индивидуальные скрещивания с

особями линии w^{1118}, их потомство отбирали на наличие встройки по селективному признаку – красным глазам (DsRed). Последующее вырезание селективного маркера мы проводили путем скрещивания линии, несущей встройку части донорного вектора, с линией, экспрессирующей PB транспозазу в эмбриогенезе (линия #8285 из Bloomington Drosophila Stock Center, Блумингтон, США).

Результаты: Коммерческий вектор pHD-ScarlessDsRed имеет линкерные последовательности с набором сайтов рестрикции для встраивания плеч для гомологичной рекомбинации с обеих сторон от селективного маркера *dsRed*. При вырезании *dsRed* PB транспозазой эти линкерные последовательности остаются в геноме. Таким образом, помимо вносимой в ген мутации, при использовании коммерческого вектора в гене остаются дополнительные последовательности, которые могут оказывать влияние на его экспрессию. Мы решили доработать коммерческий вектор, для чего соединение селективного маркера *dsRed* с плечами проводили не путем лигирования ДНК по сайтам рестрикции в линкерах, фланкирующих *dsRed*, а путем сшивающей ПЦР. Сначала отдельно амплифицировались два гомологичных плеча и *dsRed* с использованием праймеров, перекрывающих каждое из плеч и *dsRed* в местах их стыка. Затем проводили сшивающую ПЦР с получением одного общего фрагмента с использованием двух внешних праймеров. Такой подход позволил удалить линкерные последовательности из коммерческого вектора pHD-ScarlessDsRed.

В последующей работе мы столкнулись со сложностями при вырезании селективного маркера PB транспозазой. При анализе нескольких сотен потомков ни одного случая вырезания *dsRed* не произошло, что свидетельствует о низкой эффективности работы PB транспозазы в нашем случае. Мы провели второе скрещивание между особями первого поколения, несущих встройку донорного вектора и PB транспозазу. Во втором поколении мы получили несколько особей, у которых отсутствовал селективный маркер *dsRed* и присутствовала точечная замена в кодирующей части гена *orb*.

Выводы: Мы усовершенствовали коммерческий вектор pHD-ScarlessDsRed путем использования сшивающей ПЦР для соединения гомологичных плеч с селективным маркером, что позволило получить линию мух, не имеющую следов линкеров при вырезании *dsRed* из гена *orb*. Этот подход может быть использован не только для получения линий с точечными мутациями в других генах, но и для других типов мутаций, таких как делеции и инсерции.

Финансирование: Работа поддержана Российским научным фондом (№ 24-24-00517).

The use of CRISPR/Cas9 scarless system to generate *Drosophila melanogaster* line with point mutations

Zhukova M.[1]*, Yakovlev K.[2], Shidlovskii Y.[1, 3]

[1] *Institute of Gene Biology, RAS, Moscow, Russia*

[2] *A.V. Zhirmunsky National Scientific Center of Marine biology, FEB RAS, Vladivostok, Russia*

[3] *I.M. Sechenov First Moscow State Medical University, Moscow, Russia*

* *zhukovamv@gmail.com*

Key words: site-directed mutagenesis; CRISPR/Cas9; Drosophila; genetic engineering

Motivation and Aim: Our goal was to obtain *Drosophila melanogaster* line with a single nucleotide substitution in the *orb* gene, leading to the appearance of a stop codon and synthesis of the truncated protein. This mutation is required for further study of the functional role of the non-conserved C-terminal sequence of Orb.
Methods and Algorithms: The CRISPR/Cas9 scarless genome editing system is based on the traceless excision of selective markers from the integration site using piggyBac (PB) transposase. PB transposons target TTAA sites in the genome for insertion. TTAA sites are duplicated during transposon insertion and remain in a single copy after transposon excision by PB transposase. In the commercial donor vector pHD-ScarlessDsRed for the scarless editing system, the selective marker *dsRed* is flanked by PB transposon ends and TTAA sites to ensure its removal by PB transposase. When selecting a cleavage site for Cas9-endonuclease in a genome, it is important that the site is in close proximity to the TTAA insertion site. In this case the sequences of the donor vector and the genomic sequence will be different and Cas9-endonuclease will excise only the genomic DNA and not the donor vector. We selected Cas9 cleavage site 252 nt away from the point mutation. We used standard methods of genetic engineering, such as PCR and molecular cloning. To obtain a point mutation, we introduced a corresponding substitution into one of the two arms surrounding the selective marker *dsRed* in the donor vector. The point substitution was introduced into two overlapping complementary primers, and two PCR products were amplified with these primers. The resulting products were mixed and stitching PCR was performed using primers complimentary to the 5' and 3' ends of the total fragment. An aqueous DNA solution containing the donor vector and a plasmid expressing sgRNA was injected into early embryos of a fly line expressing Cas9 endonuclease (line #55821 from Bloomington Drosophila Stock Center, Bloomington, USA). Flies developed from embryos were individually crossed with individuals of line w^{1118}, their offspring were selected for the presence of the insertion by the selective trait – red eyes (DsRed). Subsequently we performed excision of the selective marker by crossing the line carrying the insertion part of the donor vector with a line expressing PB transposase in embryogenesis (line #8285 from Bloomington Drosophila Stock Center, Bloomington, USA).
Results: The commercial vector pHD-ScarlessDsRed has linker sequences with a set of restriction sites to insert shoulders for homologous recombination on both sides of the selective marker *dsRed*. When *dsRed* is excised by PB transposase, these linker sequences remain in the genome. Thus, in addition to the point mutation in the gene, when using a commercial vector, additional sequences remain that may affect gene expression. We decided to improve the commercial vector, for which the connection of the selective marker *dsRed* to the shoulders was carried out by stitching PCR rather than by ligating DNA at restriction sites in linkers flanking *dsRed*. First, two homologous shoulders and *dsRed* were amplified separately using primers that overlapped each of the shoulders and *dsRed* at their junction sites. Stitching PCR was then performed to obtain one common fragment using two external primers. This approach allowed us to remove linker sequences from the commercial vector pHD-ScarlessDsRed.
In subsequent work, we encountered difficulties in excision of the selective marker by PB transposase. When we analyzed several hundred offspring, there was no a single case of excision of *dsRed*, indicating the low efficiency of PB transposase in our case. We performed a second cross between individuals of the first generation carrying the donor vector insert and PB transposase. In the second generation, we obtained several

individuals that lacked the selective marker *dsRed* and had a point substitution in the coding part of the *orb* gene.
Conclusion: We have improved the commercial pHD-ScarlessDsRed vector by using stitching PCR to connect homologous arms with a selective marker. This allowed us to obtain a fly line with no linker traces when *dsRed* is excised from the *orb* gene. This approach can be used not only to obtain lines with point mutations in other genes, but also for other types of mutations, such as deletions and insertions.
Funding: The work is supported by the Russian Science Foundation (No. 24-24-00517).

doi 10.18699/bgrs2024-7.1-04

Поиск маркеров, пригодных для цитогенетической идентификации отдельных хромосом, в кариотипе живородящей ящерицы *Zootoca vivipara*

Жукова Ю.[1]*, Куприянова Л.[2], Попова М.[3], Галкина С.[1]

[1] *Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*

[2] *Зоологический институт РАН, Санкт-Петербург, Россия*

[3] *Сколковский институт науки и технологий, Москва, Россия*

** st121062@student.spbu.ru*

Ключевые слова: Reptilia; кариотип; теломеры; флуоресцентная гибридизация *in situ* (FISH); хромосомная изменчивость

Мотивация и цель: *Zootoca vivipara* (Ящерица живородящая) является широко распространенным видом семейства Lacertidae (Reptilia). Этот вид обитает почти на всей территории Северной, Центральной и Восточной Европы, а также на севере Азии. Данный вид ящериц интересен своей хромосомной изменчивостью. На основании исследований кариотипов было показано, что вид *Z. vivipara* включает морфологически слабо дифференцированные подвиды и хромосомные формы с неопределенным таксономическим статусом [1]. К настоящему времени в пределах данного вида обнаружено 7 хромосомных форм [2]. Объектом нашего исследования стала восточная (русская) форма живородящей ящерицы, обитающая в восточной части Европы. Ее кариотип характеризуется разным числом хромосом у самцов ($2n = 36$) и у самок ($2n = 35$). В хромосомный набор входят 32 аутосомы и различное количество половых хромосом: 2 пары Z-хромосом у самцов ($Z_1Z_1Z_2Z_2$) и Z_1Z_2W у самок [3]. Кариотип очень сложен с цитогенетической точки зрения: все хромосомы русской формы, включая половые, являются акроцентриками, что значительно усложняет их идентификацию. С помощью С-бэндинга возможно отличить только W-хромосому от аутосом, поскольку она обладает заметными центромерным, теломерным и интерстициальным бэндами [4]. Помимо митотических хромосом, в литературе присутствуют данные и о хромосомах, полученных из сперматоцитов на стадии диакинеза профазы I мейоза ($n = 18$) [5]. На сегодняшний день в общедоступной базе генетических данных Национального центра биотехнологической информации США (NCBI) имеется три сборки расшифрованного генома живородящей ящерицы, две из них собраны до хромосомного уровня [6]. Эти данные могут быть использованы для поиска многокопийных последовательностей, специфичных для отдельных хромосом и пригодных для разработки молекулярных зондов. Описанные ранее рассеянные повторы SINE-Zv 700, SINE-Zv 300 [7], Zv516 и Zv817 [8] локализуются практически на всех хромосомах и не могут быть использованы в качестве таких маркеров. Цель нашей работы – поиск хромосом-специфичных маркеров и разработка зондов, пригодных для идентификации в кариотипе *Zootoca vivipara* отдельных хромосом, в том числе половых, методом флуоресцентной гибридизации *in situ* (FISH).

Методы и алгоритмы: Для получения препаратов хромосом живородящей ящерицы были получены суспензии фиксированных клеток из красного костного мозга. Ящерицы были пойманы в дикой природе на территории Ленинградской области (поселок Вырица). Цитогенетическое картирование хромосом проводили при помощи FISH с зондами к тандемным повторам, включая теломерные, биоинформатически рассчитанные повторы Zviv84A, Zviv84B и тандемные повторы, содержащие мотив CenpB. Для идентификации данных повторов в геноме ящерицы использовали специальный скрипт Python и специальный C++ скрипт [9]. К данным повторам были разработаны прямомеченные зонды Zviv84A Cy3 (5'-CAYACAGGGGAGAAACCMTWTMARTGYWTGGARTGTGGAAAG) и Zviv84B FAM (5'-SCRMAGRTCGGCGGGCGCTGTTTGCCGGGCTTGTCAAGC CMRSMWMV). Для поиска последовательностей 5S рибосомной ДНК (рДНК) и генов ядрышкового организатора в сборке генома живородящей ящерицы использовали последовательности соответствующих генов рРНК домашней курицы (GenBank X01309.1, KT445934.2). Поиск осуществляли с помощью алгоритма NCBI BLAST (https://blast.ncbi.nlm.nih.gov/). Анализ последовательностей рДНК осуществляли в программе Geneious Prime 2023.0.1 (Biomatters Ltd.). В качестве зондов для генов рРНК использовали олигонуклеотид к гену 5S рРНК (5'-ACAGCACCCGGTATTCCCAGGCGGTCTCCCAT CCAAGTACTAACCGGG) и плазмиду, содержащую фрагмент гена 18S рРНК японского перепела, на 99 % гомологичную последовательности этого же гена у живородящей ящерицы.
Результаты: При картировании зондов на митотических хромосомах ящерицы нами обнаружено, что блоки теломерного повтора (TTAGGG)n находятся только на концах хромосом, интерстициальных сайтов не выявлено. Сходное распределение наблюдалось и при картировании на хромосомах-ламповых щетках. Полученный результат совпадает с имеющимися литературными данными о том, что интерстициальные теломерные сайты довольно редко встречаются в кариотипах чешуйчатых рептилий [10]. Zviv84A в геномной сборке сконцентрирован на 2, 5, 9, 10 и 13 хромосомах. Zviv84B рассеян по хромосомам. При гибридизации с Zviv84A обнаружены яркие прителомерные сигналы только на хромосомах 2 и 5. Гибридизация с Zviv84B дает рассеянные сигналы на хромосомах, поэтому этот повтор не может быть использован в качестве хромосомного маркера. Кластеры генов 5S и 18S рРНК располагаются в геноме на хромосомах 5 и 17 соответственно.
Выводы: Зонды к тандемному повтору Zviv84A и кластерам генов 5S и 18S рРНК могут быть использованы для идентификации хромосом 2, 5 и 17 в кариотипе *Zootoca vivipara.*
Благодарность: Авторы выражают благодарность РЦ ЦКП «Хромас» Научного парка Санкт-Петербургского государственного университета.

Search for markers suitable for cytogenetic identification of individual chromosomes in the karyotype of the common lizard *Zootoca vivipara*

Zhukova J.[1]*, Kupriyanova L.[2], Popova M.[3], Galkina S.[1]

[1] *Saint Petersburg State University, St. Petersburg, Russia*

[2] *Zoological Institute, RAS, St. Petersburg, Russia*

[3] *Skolkovo Institute of Science and Technology, Moscow, Russia*

* *st121062@student.spbu.ru*

Key words: Reptilia; karyotype; telomeres; fluorescence *in situ* hybridization (FISH); chromosomal variation

Motivation and Aim: *Zootoca vivipara* (the common lizard) is a widespread species of the family Lacertidae (Reptilia). This species is native to much of northern Eurasia, including northern, central and eastern Europe, as well as northern Asia. This reptile is interesting for its chromosomal variability. Based on karyotype studies, it was shown that the species *Z. vivipara* includes morphologically poorly differentiated subspecies and chromosomal forms with uncertain taxonomic status [1]. To date, 7 chromosomal forms have been found within this species [2]. The object of our study was the eastern (Russian) form of the viviparous lizard inhabiting the eastern part of Europe. Its karyotype is characterized by a different number of chromosomes in males ($2n = 36$) and females ($2n = 35$). The chromosome set includes 32 autosomes and a varying number of sex chromosomes: two pairs of Z chromosomes in males ($Z_1Z_1Z_2Z_2$) and Z_1Z_2W in females [3]. The cytogenetic description of the karyotype is challenging because all chromosomes of the Russian form, including the sex chromosomes, are acrocentric. This morphology significantly complicates chromosome identification. Using C-banding, only the W chromosome ca be distinguished from autosomes due to its prominent centromeric, telomeric, and interstitial bands [4]. In addition to mitotic chromosomes, data on meiotic chromosomes from spermatocytes at the diakinesis stage ($n = 18$) are also available [5]. To date, the publicly available genetic database of the US National Centre for Biotechnology Information (NCBI) contains threeassemblies of the decoded genome of the viviparous lizard, two of which are assembled to the chromosomal level [6]. These data can be used to search for multicopy sequences specific to individual chromosomes, suitable for developing molecular probes. The previously described interspersed repeats SINE-Zv 700, SINE-Zv 300 [7], Zv516 and Zv817 [8] are localized on almost all chromosomes and are therefore unsuitable as such markers. The aim of our work is to search for chromosome-specific markers and develop probes suitable for the identification of individual chromosomes, including sex chromosomes, in the karyotype of *Zootoca vivipara* by fluorescence *in situ* hybridization (FISH).

Methods and Algorithms: Chromosome preparations were obtained from fixed red bone marrow. Lizards were caught in the wild on the territory of the Leningrad region (in Vyritsa village). We performed FISH with probes to tandem repeats, including telomeric, bioinformatically calculated repeats Zviv84A, Zviv84B and repeats containing the CenpB motif. A special Python script and a special C++ script were used to identify these repeats in the lizard genome [9]. Directly labelled probes Zviv84A Cy3 (5'-CAYACAGGGGAGAAACCMTWTMARTGYWTGGARTGTGGAAAG) и Zviv84B FAM (5'-SCRMAGRTCGGCGGGCGCTGTTTGCCGGGCTTGT CAAGCCMRSMWMV) were designed. A domestic chicken rRNA sequence (GenBank X01309.1) was used to search for 5S ribosomal DNA (rDNA) sequences and core organizer genes in the live-bearing lizard genome assembly. Searches were performed using the NCBI BLAST algorithm (https://blast.ncbi.nlm.nih.gov/). The rDNA sequences were analyzed using Geneious Prime 2023.0.1 software (Biomatters Ltd.). repeats containing the CenpB motive. A custom Python script and a custom C++ script were used to identify these repeats in the lizard genome [9]. Directly labelled probes Zviv84A Cy3 (5'-CAYACAGGGGAGAAACCMTWTMARTGYWTGGARTGTGG AAAG) и Zviv84B FAM (5'-SCRMAGRTCGGCGGGCGCTGTTTGCCGGGCTTGT CAAGCCMRSMWMV) were designed for these repeats. To search for 5S ribosomal

DNA (rDNA) sequences and nucleolar organizer genes in the lizard genome assembly, we used the corresponding domestic chicken rRNA genes sequences (GenBank X01309.1, KT445934.2). Searches were performed using the NCBI BLAST algorithm (https://blast.ncbi.nlm.nih.gov/). The rDNA sequences were analyzed using Geneious Prime 2023.0.1 software (Biomatters Ltd.).

Results: We found that telomeric repeat (TTAGGG)n blocks are only located at chromosome ends, no specific signals in interstitial sites were observed on mitotic chromosomes after FISH. A similar distribution was observed when high resolution mapping on chromosomes at the lampbrush stage. This result coincides with data reported in the literature that interstitial telomeric sites are rather rare in karyotypes of Squamata [10]. Zviv84A in the genomic assembly is concentrated on chromosomes 2, 5, 9, 10 and 13. Zviv84B is dispersed on chromosomes. FISH with the Zviv84A probe revealed bright peritelomeric signals on chromosomes 2 and 5. FISH with the Zviv84B probe gives interspersed signals, so this repeat is unsuitable as a chromosomal marker. The 5S and 18S rRNA gene clusters are located in the genome on chromosomes 5 and 17 respectively.

Conclusion: Probes specific to the tandem repeat Zviv84A and the 5S and 18S rRNA gene clusters can be used to identify chromosomes 2, 5 and 17 in the karyotype of *Zootoca vivipara.*

Acknowledgments: The authors are grateful to the Research Resource Center “Chromas” of the Science Park of St. Petersburg State University.

Список литературы/References

1. Куприянова Л.А., Мелащенко О.Б. Новые данные о внутривидовом кариотипическом разнообразии живородящей ящерицы (*Zootoca vivipara* (Lichtenstein 1823), Lacertidae): распространение и расселение на крайнем западе России. *Зоологический журнал*. 2015;1(94):106-110
   [Kupriyanova L.A., Melashchenko O.B. New data on intraspecific karyotypic diversity of a viviparous lizard (*Zootoca vivipara* (Lichtenstein 1823), Lacertidae): distribution and settlement in the far west of Russia. *Zoological J*. 2015;1(94):106-110 (in Russian)]
2. Surget-Groba Y. et al. Intraspecific phylogeography of *Lacerta vivipara* and the evolution of Viviparity. *Mol Phylogenet Evol*. 2001;3(18):449-459
3. Safronova L.D., Kupriyanova L.A. Metaphase and meiotic chromosomes, synaptonemal complexes (SC) of the lizard Zootoca vivipara. *Russ J Genet*. 2016;11(52):1186-1191
4. Kupriyanova L.A. Karyotype diversity of the Eurasian lizard *Zootoca vivipara* (Jacquin, 1787) from Central Europe and the evolution of viviparity. In: Proceedings of the 13th Congress of the Societas Europaea Herpetologica. 2006;67-72
5. Kupriyanova L., Safronova L. A brief review of meiotic chromosomes in early spermatogenesis and oogenesis and mitotic chromosomes in the viviparous lizard *Zootoca vivipara* (Squamata: Lacertidae) with multiple sex chromosomes. *Animals*. 2022;1(12):19
6. www.ncbi.nlm.nih.gov/datasets/genome/?taxon=8524
7. Petraccioli A., Guarino F.M., Kupriyanova L. et al. Isolation and characterization of interspersed repeated sequences in the common lizard, *Zootoca vivipara*, and their conservation in squamata. *Cytogenet Genome Res*. 2019;157(1-2):65-76
8. Mezzasalma M., Capriglione T., Kupriyanova L. et al. Characterization of two transposable elements and an ultra-conserved element isolated in the genome of *Zootoca vivipara* (Squamata, Lacertidae). *Life*. 2023;13:637
9. Попова М.А. Анализ разнообразия больших тандемных повторов в геноме хищников. Выпускная квалификационная работа магистра. 2023
   [Popova M.A. Analysis of the Diversity of Large Tandem Repeats in Carnivora. Graduation thesis. 2023 (in Russia)]
10. Rovatsos M. et al. Interstitial telomeric motifs in squamate reptiles: when the exceptions outnumber the rule. *PLoS One*. 2015;8(10):e0134985

doi 10.18699/bgrs2024-7.1-05

Описание кариотипа и сравнительный анализ генома глухаря обыкновенного (*Tetrao urogallus*)

Иванова Е.[1, 2]*, Беклемишева В.[1], Галкина С.[3], О'Коннор Б.[4], Гриффин Д.[4], Графодатский А.[1], Проскурякова А.[1]

[1] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[3] *Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*

[4] *Университет Кента, Кентербери, Великобритания*

* *ekater1na.s.ivanova@yandex.ru*

Ключевые слова: *Gallus gallus*; *Burhinus oedicnemus*; сравнительный хромосомный пэйнтинг; бактериальные искусственные хромосомы (BAC); конститутивный гетерохроматин; рибосомные гены; теломера

Мотивация и цель: Обыкновенный глухарь (*Tetrao urogallus*) является распространенным видом на территории Европы и Азии, в том числе занимает широкий ареал на территории России. Глухарь входит в состав отряда курообразные, семейство фазановые (Galliformes, Phasianidae). Отряд курообразные входит в базальную ветвь клады новонебных (Neognathae), охватывающей большинство современных птиц. Несмотря на наличие Hi-C сборки генома хромосомного уровня, выполненной в рамках проекта Darwin Tree of Life [1], в литературе отсутствуют данные о кариотипе обыкновенного глухаря. Целью настоящего исследования стало всестороннее изучение хромосомного набора глухаря и сравнение его с кариотипом гипотетического общего предка новонебных (Neognathae). Для достижения этой цели необходимо описать кариотип глухаря с применением различных методов дифференциального окрашивания хромосом, локализовать методом FISH уникальные последовательности курицы (*Gallus gallus*), клонированные в искусственных бактериальных хромосомах (BAC), и хромосомоспецифичные зонды авдотки (*Burhinus oedicnemus*), а также привязать хромосомные скаффолды к хромосомам глухаря.

Методы и алгоритмы: Для получения препаратов хромосом обыкновенного глухаря получены первичные клеточные культуры фибробластов и суспензии фиксированных метафазных клеток по ранее разработанной методике [2]. Для описания кариотипа с использованием красителя Гимза получены рутинно окрашенные хромосомы и применены различные методы дифференциального окрашивания (GTG-, CBG-, CDAG-) [3–5]. Для выявления гомологичных районов в кариотипах обыкновенного глухаря и авдотки методом флуоресцентной *in situ* гибридизации (FISH) [6] были использованы хромосомоспецифичные зонды, разработанные для авдотки [7]. Для сравнения кариотипов глухаря и курицы были использованы BAC-клоны, содержащие уникальные последовательности курицы (CHORI-261) [8]. Для локализации теломерных и ядрышкообразующих районов в кариотипе глухаря были использованы пробы, содержащие теломерные повторы [9] и зонды к генам 18S и 28S рРНК [10]. Все зонды были синтезированы с включением меченых нуклеотидов и локализованы на хромосомах. Выравнивание

сборок геномов хромосомного уровня курицы [11] и обыкновенного глухаря [1] и визуализация синтенных групп были осуществлены в программе Geneious.
Результаты: Впервые описан кариотип обыкновенного глухаря с помощью рутинного окрашивания и методов GTG-, CBG-, CDAG-бэндинга. Наши результаты по подсчету диплоидного числа хромосом у этого вида ($2n$ = 78) совпадают с более ранними оценками и в целом согласуются с количеством скаффолдов, полученных при сборке генома. На основании данных FISH с пробами хромосом авдотки и BAC-клонов курицы, а также с теломерными повторами и рибосомными генами были выявлены ортологичные районы на 29 аутосомах и Z-хромосоме обыкновенного глухаря. На основании этих данных стало возможным соотнести хромосомные скаффолды с хромосомами обыкновенного глухаря. Нами были обнаружены несоответствия между физическим размером хромосом и относительным размером скаффолдов.
Показано согласование молекулярно-цитогенетических и сравнительно-биоинформатических данных, которые выявили слияние хромосом 6 и 8 и разрыв хромосомы 2, ортологичной хромосомам курицы, в кариотипе глухаря. Ортолог хромосомы 4 курицы сохраняет структуру общего предка новонебных, представленную хромосомой 4 и микрохромосомой. Выравнивание геномов в Geneious показало, что макрохромосомы глухаря сохраняют порядок консервативных сегментов, характерный для курицы. Локализация BAC-клонов курицы на хромосомах глухаря с помощью FISH подтвердила данные сравнения геномов, полученные при анализе *in silico*. В микрохромосомах не было выявлено транслокаций как биоинформатически, так и с помощью FISH. Инверсии на хромосомах глухаря 6, 10, 17, 23 и 28 выявлены с помощью биоинформатического анализа, но не подтвердились молекулярно-цитогенетическими данными.
Выводы: Впервые получено полное описание кариотипа обыкновенного глухаря (*Tetrao urogallus,* $2n$ = 78) с использованием методов классической, дифференциальной и молекулярной цитогенетики. В результате интеграции биоинформатического и цитогенетического подходов установлено соответствие между хромосомными скаффолдами и хромосомами обыкновенного глухаря. Согласно полученным данным, кариотип глухаря высококонсервативен и практически не отличается от кариотипа общего предка новонебных (Neognathae) [12]. Хромосомы 6, 10, 17, 23 и 28 обыкновенного глухаря характеризуются наличием инверсий в сравнении с ортологичными хромосомами курицы. Было показано, что кариотип обыкновенного глухаря отличается от кариотипа курицы двумя слияниями (ортологичными хромосомам 6 и 8, 31 и 35 курицы) и разрывом элемента, ортологичного хромосоме 2 курицы.
Финансирование: Исследование поддержано грантом РНФ (№ 23-74-01016).

Description of the karyotype and comparative analysis of the western capercaillie (*Tetrao urogallus*) genome

Ivanova E.[1, 2]*, Beklemisheva V.[1], Galkina S.[3], O'Connor R.[4], Griffin D.K.[4], Graphodatsky A.[1], Proskuryakova A.[1]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Saint Petersburg State University, St. Petersburg, Russia*

[4] *University of Kent, Canterbury, UK*

* *ekater1na.ivanova@live.ru*

Key words: *Gallus gallus; Burhinus oedicnemus*; comparative chromosome painting; bacterial artificial chromosomes (BAC); constitutive heterochromatin; ribosomal genes; telomere

Motivation and Aim: Western capercaillie (*Tetrao urogallus*) is a common species in Europe and Asia, including the territory of Russia. The capercaillie is a part of order Galliformes, family Phasianidae. Galliformes belong to the basal branch of Neognathae, which includes the most modern birds. Despite the presence of a chromosome level genome assembly of capercaillie based on Hi-C technologies, which was obtained as part of the Darwin Tree of Life project [1], there is no published data on the karyotype of western capercaillie. The purpose of this study was to conduct a comprehensive analysis of the chromosomal set of western capercaillie and to compare it with the karyotype of the hypothetical common ancestor of neopalates (Neognathae). To achieve this goal, it was necessary to characterize the karyotype of western capercaillie using various methods of differential chromosome staining, localize the set of chicken (*Gallus gallus*) sequences cloned in bacterial artificial chromosomes (BAC) and chromosome-specific probes of Eurasian stone-curlew (*Burhinus oedicnemus*) using FISH, and link chromosome scaffolds the chromosomes.
Methods and Algorithms: To obtain metaphase plates of western capercaillie, cell cultures of fibroblasts and suspensions of fixed metaphase cells were obtained according to a previously developed method [2]. The karyotype was described using Giemsa staining, routine staining and various differential staining methods (GTG-, CBG-, CDAG-) [3–5]. To identify homologous regions in the karyotypes of western capercaillie and Eurasian stone-curlew with fluorescent in situ hybridization (FISH) [6], chromosome-specific probes developed for Eurasian stone-curlew were used [7]. To compare the karyotypes of capercaillie and chicken, BAC clones containing unique chicken sequences (CHORI-261) were used [8]. To localize telomeric and nucleolus organizer regions in western capercaillie karyotype, samples containing telomeric repeats [9] and probes for the 18S and 28S rRNA genes [10] were used. All probes were synthesized with labeled nucleotides and localized on chromosomes. Alignment of chromosome level genome assemblies of chicken [11] and capercaillie [1] and visualization of syntenic groups was carried out in Geneious.
Results: For the first time, the karyotype of the western capercaillie was described using routine staining and GTG-, CBG-, and CDAG-banding methods. Our results of counting the diploid number of chromosomes in this species ($2n = 78$) coincide with earlier estimates and are generally consistent with the number of scaffolds obtained during genome assembly. Based on FISH data with samples of Eurasian stone-curlew chromosomes and chicken BAC-clones, as well as with telomeric repeats and ribosomal genes, orthologous regions were identified on 29 autosomes and the Z-chromosome of western capercaillie. Based on these data, it became possible to correlate chromosome scaffolds with chromosomes of western capercaillie. We found discrepancies between the physical size of chromosomes and the relative size of scaffolds.
The agreement between molecular cytogenetic and comparative bioinformatics data is shown, which revealed the fusion of chromosomes 6 and 8 and the fission of chromosome 2 orthologous to chicken chromosomes, in the karyotype of capercaillie. The ortholog of chicken chromosome 4 in capercaillie retains structure of the common ancestor of neopalates being presented by chromosome 4 and microchromosome.

Genome alignment in Geneious showed that the capercaillie macrochromosomes retain the order of conserved segments of the chicken. Localization of chicken BAC clones on western capercaillie chromosomes using FISH confirmed the genome comparison data obtained from *in-silico* analysis. No translocations were detected in microchromosomes either bioinformatically or using FISH. Inversions on the 6, 10, 17, 23 and 28 chromosomes of the capercaillie were identified using bioinformatics analysis, but were not supported by molecular cytogenetic data.

Conclusions: For the first time, a complete description of the karyotype of western capercaillie (*Tetrao urogallus*, $2n = 78$) has been obtained using the methods of classical, differential and molecular cytogenetics. As a result of bioinformatic and cytogenetic approaches, correspondences were established between the scaffolds and chromosomes of western capercaillie. According to the data obtained, the karyotype of capercaillie is highly conservative and highly similar with the karyotype of the common ancestor of neopalates (Neognathae) [12]. Chromosomes 6, 10, 17, 23 and 28 of western capercaillie differ from the orthologous chromosomes of the chicken by inversions. It was shown that the karyotype of the common capercaillie differs from the chicken karyotype by two fusions (orthologous to chicken chromosomes 6 and 8, 31 and 35) and a fission of an element orthologous to chicken chromosome 2.

Funding: The study is supported by the research grant No. 23-74-01016 from the Russian Science Foundation.

Список литературы/References

1. https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_951394365.1/
2. Romanenko S.A. et al. Segmental paleotetraploidy revealed in sterlet (Acipenser ruthenus) genome by chromosome painting. *Mol Cytogenet.* 2015;8(1):90
3. Seabright M. A rapid banding technique for human chromosomes. *Lancet*. 1971;298(7731):971-972
4. Sumner A.T. A simple technique for demonstrating centromeric heterochromatin. *Exp Cell Res.* 1972;75(1):304-306
5. Lemskaya N.A. et al. A combined banding method that allows the reliable identification of chromosomes as well as differentiation of AT-and GC-rich heterochromatin. *Chromosom Res.* 2018;12(26):307-315
6. Nie W. et al. Avian comparative genomics: Reciprocal chromosome painting between domestic chicken (Gallus gallus) and the stone curlew (Burhinus oedicnemus, Charadriiformes). An atypical species with low diploid number. *Chromosom Res.* 2009;17(1):99-113
7. https://bacpacresources.org
8. Liehr T. Fluorescence in Situ Hybridization (FISH). Springer, 2017
9. Maden B.E.H. et al. Clones of human ribosomal DNA containing the complete 18 S-rRNA and 28 S-rRNA genes. Characterization, a detailed map of the human ribosomal transcription unit and diversity among clones. *Biochem J*. 1987;246(2):519-527
10. Ijdo J.W. et al. Improved telomere detection using a telomere repeat probe (TTAGGG) n generated by PCR. *Nucleic Acids Res*. 1991;19(17):4780
11. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_016699485.2/
12. Kretschmer R. et al. Interspecies chromosome mapping in caprimulgiformes, piciformes, suliformes, and trogoniformes (Aves): Cytogenomic insight into microchromosome organization and karyotype evolution in birds. *Cells.* 2021;10(4):826

doi 10.18699/bgrs2024-7.1-06

Кариотип хохлатого оленя (*Elaphodus cephalophus*): описание и сравнительная геномика

Иванова Е.[1, 2]*, Янг Ф.[3], Фергюсон-Смит М.[4], Ларкин Д.[5], Графодатский А.[1], Проскурякова А.[1]

[1] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[3] *Школа медицины и наук о жизни, Шандонский технологический университет, Цзыбо, Китай*

[4] *Кембриджский ресурсный центр сравнительной геномики, Кембриджский университет, Кембридж, Великобритания*

[5] *Королевский ветеринарный колледж, Лондон, Великобритания*

* *ekater1na.s.ivanova@yandex.ru*

Ключевые слова: *Camelus dromedarius*; *Bos taurus*; сравнительный хромосомный пэйнтинг; бактериальные искусственные хромосомы (BAC); конститутивный гетерохроматин; рибосомные гены; теломера

Мотивация и цель: Хохлатый олень (*Elaphodus cephalophus*) входит в состав семейства оленевых (Cervidae) и является единственным представителем рода *Elaphodus*. Филогенетическое положение вида до сих пор остается предметом дискуссий, однако, согласно последним исследованиям, род *Elaphodus* является сестринским родом для *Muntiacus* и относится к одной монофилетической группе с мунтжаками (Muntiacini) [1]. Триба Muntiacini – одна из интереснейших с точки зрения цитогенетики за счет многочисленных слияний хромосом в кариотипах этих видов. Несмотря на пристальное внимание исследователей к мунтжакам, наличие большого количества цитогенетических и геномных исследований, хохлатый олень – вид, признанный находящимся под угрозой исчезновения, мало изучен относительно других представителей своей трибы. Для хохлатого оленя не получена качественная сборка генома, нет подробной хромосомной карты, на основе которой можно было бы оценить внутрихромосомные перестройки и провести сравнение с другими видами оленей. Таким образом, целью нашей работы стало получение как можно более полных данных об организации кариотипа хохлатого оленя.

Методы и алгоритмы: Клеточные культуры фибробластов для получения суспензий метафазных хромосом [2] были предоставлены Кембриджским ресурсным центром. Описание кариотипа было осуществлено с применением рутинной окраски хромосом Гимзой и различных методов бэндинг цитогенетики (GTG, CBG, CDAG) [3–5]. Для сравнения кариотипа хохлатого оленя с кариотипами других видов жвачных были локализованы зонды, содержащие BAC-клоны коровы (*Bos taurus*) (CHORI-240) [6] и сортированные хромосомы верблюда (*Camelus dromedarius*) [7], методом флуоресцентной *in situ* гибридизации (FISH) [8].

Результаты: Анализ кариотипа хохлатого оленя ($2n = 48$) выявил наличие нео-X и нео-Y хромосом. Такой вариант кариотипа уже был ранее описан [9]. Были локализованы 75 BAC-клонов, соответствующих 29 аутосомам коровы и 29 BAC-клонов с X-хромосомы коровы на хромосомах хохлатого оленя. В ранее

опубликованной работе [10] было выявлено наличие горячих точек эволюции в кариотипе черного мунтжака (*Muntiacus crinifrons*) на хромосомах, гомологичных 3, 22, 10, 33, 24, 30 хромосомам верблюда: 3/22/3/22, 33/10, 24/30. Для проверки наличия перестроек в этих горячих точках у хохлатого оленя выборочно были локализованы зонды, содержащие вышеперечисленные сортированные хромосомы верблюда. Наличие двух из трех инверсий (3/22/3/22/3/22 и 24/30/24) в кариотипе хохлатого оленя подтвердилось как с помощью пэйнтинг-проб верблюда, так и данными локализации BAC-клонов. CDAG окрашивание хромосом хохлатого оленя выявило накопление гетерохроматина с образованием крупных блоков на X- и Y-хромосомах.

Согласно полученным данным, кариотип хохлатого оленя образовался за счет тандемных слияний, аутосомных транслокаций на половые хромосомы, инверсий и экспансии гетерохроматина. Локализация BAC-клонов с X-хромосомы коровы показала, что X-хромосома хохлатого оленя обладает схожим порядком консервативных блоков с другими представителями настоящих оленей (Cervinae), у общего предка которых произошли одна инверсия и изменение положения центромеры. Как и у других представителей трибы мунтжаков [11], у хохлатого оленя произошла транслокация X-хромосомы на аутосому с образованием нео-X-хромосомы.

Выводы: В результате работы была получена первая сравнительная хромосомная карта хохлатого оленя с BAC-клонами коровы. Порядок BAC-клонов показал, что кариотип хохлатого оленя образовался за счет тандемных слияний. С помощью FISH был показан разрыв предкового элемента, соответствующего хромосоме 7 коровы, и слияние с p-плечом хромосомы 7 хохлатого оленя. Также показана инверсия на хромосоме 3 хохлатого оленя фрагмента, соответствующего хромосоме 24 коровы. В ходе эволюции X-хромосомы у хохлатого оленя произошли аутосомная транслокация и накопление крупного блока гетерохроматина, однако X-хромосома сохраняет порядок консервативных сегментов, характерный для настоящих оленей.

Финансирование: Исследование поддержано грантом РНФ (19-14-00034) и грантом программы фундаментальных научных исследований FWGZ-2021-0015 (122011800125-7).

Karyotype of tufted deer (*Elaphodus cephalophus*): description and comparative genomics

Ivanova E.[1, 2]*, Yang F.[3], Ferguson-Smith M.[4], Larkin D.[5], Graphodatsky A.[1], Proskuryakova A.[1]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *School of Medicine and Life Sciences, Shandong University of Technology, Zibo, China*

[4] *Cambridge Resource Centre for Comparative Genomics, University of Cambridge, Cambridge, UK*

[5] *Royal Veterinary College, London, UK*

* *ekater1na.s.ivanova@yandex.ru*

Motivation and Aim: Tufted deer (*Elaphodus cephalophus*) is a part of the deer family (Cervidae) and is the only member of genus Elaphodus. The phylogenetic position of the species is still a matter of debate, however, according to recent studies, the genus

Elaphodus is considered to be a sister genus to *Muntiacus*, and belongs to the same monophyletic group with muntjacs (Muntiacini) [1]. The tribe Muntiacini is one of the most interesting for cytogenetics due to numerous chromosome fusions in the karyotypes of these species. Despite the close attention of researchers to muntjacs and a large number of cytogenetic and genomic studies, the tufted deer, a species recognized as endangered, has been studied relatively superficially to other representatives of its tribe. A high-quality genome assembly has not been obtained for the tufted deer; there is no detailed chromosomal map on the basis of which intrachromosomal rearrangements could be assessed and comparisons made with other deer species. Thus, the goal of our work was to obtain the most complete data possible on the organization of tufted deer karyotype.

Methods and Algorithms: Fibroblast cell cultures to obtain suspensions of metaphase chromosomes [2] were provided by the Cambridge Resource Centre. Karyotype description was carried out using routine chromosome staining and various banding cytogenetics methods (GTG, CBG, CDAG) [3–5]. To compare the karyotype of the tufted deer with the karyotypes of other ruminant species, probes containing cattle BAC clones of (*Bos taurus*) (CHORI-240) [6] and sorted chromosomes of dromedary (*Camelus dromedarius*) [7] were localized using fluorescent *in situ* hybridization method (FISH) [8].

Results: Analysis of the karyotype of a tufted deer ($2n = 48$) revealed the presence of neo-X and neo-Y chromosomes. This variant of karyotype has already been described previously [9]. 75 BAC clones corresponding to 29 cattle autosomes and 29 BAC clones from the cattle X chromosome were localized on the tufted deer chromosomes. Previously published work [10] revealed the presence of evolutionary breakpoint regions in the black muntjac (*Muntiacus crinifrons*) karyotype on chromosomes homologous to dromedary chromosomes 3, 22, 10, 33, 24, 30: 3/22/3/22, 33/10, 24/30. To test for the presence of the rearrangements at these hotspots in tufted deer, probes containing the sorted camel chromosomes were selectively localized. The presence of two of three inversions (3/22/3/22/3/22 and 24/30/24) in the tufted deer karyotype was confirmed both by chromosome painting and by BAC clones localization. CDAG staining of tufted deer chromosomes revealed accumulation of heterochromatin with the formation of large blocks on the X and Y chromosomes.

According to obtained data, the karyotype of the tufted deer was formed due to tandem fusions, autosomal translocations to sex chromosomes, inversions and expansion of heterochromatin. Localization of cattle X chromosomal BAC clones showed that the X chromosome of the tufted deer has a similar order of conserved blocks with other representatives of Cervinae, in whose common ancestor there was one inversion and the centromere reposition. Like other representatives of the muntjac tribe [11], the tufted deer underwent a translocation of the X chromosome onto an autosome with the formation of a neo-X chromosome.

Conclusions: As a result, the first comparative chromosome map of a tufted deer with cattle BAC clones was obtained. The order of the cattle BAC clones showed that the tufted deer karyotype was formed by tandem fusions. FISH showed a break in the ancestral element corresponding to cattle chromosome 7 and a fusion with the p-arm of tufted deer chromosome 7. An inversion was also shown on the chromosome 3 of tufted deer of a fragment corresponding to the cattle chromosome 24. During the evolution of the X chromosome in tufted deer, an autosomal translocation and accumulation of a large

block of heterochromatin occurred, but the X chromosome retains the order of conservative segments characteristic of Cervinae subfamily.

Funding: The study was supported by the research grant No. 19-14-00034 from the Russian Science Foundation, the research grant FWGZ-2021-0015 (122011800125-7) from the Ministry of Science and Higher Education (Russia) via the Institute of Molecular and Cellular Biology.

Список литературы/References

1. Jiang L. et al. Sequencing and Characterization of the Complete Mitochondrial Genome of Tufted Deer, Elaphodus cephalophus (Artiodactyla: Cervidae) and its Phylogenetic Position within the Family Cervidae. *Pak J Zool*. 2023;55(5):2019
2. Romanenko S.A. et al. Segmental paleotetraploidy revealed in sterlet (Acipenser ruthenus) genome by chromosome painting. *Mol Cytogenet.* 2015;8(1):90
3. Seabright M. A rapid banding technique for human chromosomes. *Lancet*. 1971;298(7731):971-972
4. Sumner A.T. A simple technique for demonstrating centromeric heterochromatin. *Exp Cell Res.* 1972;75(1):304-306
5. Lemskaya N.A. et al. A combined banding method that allows the reliable identification of chromosomes as well as differentiation of AT-and GC-rich heterochromatin. *Chromosom. Res.* 2018; 12 (26):307-315
6. https://bacpacresources.org
7. Balmus G. et al. Cross-species chromosome painting among camel, cattle, pig and human: further insights into the putative Cetartiodactyla ancestral karyotype. *Chromosome Res*. 2007;15:499-514
8. Liehr T. Fluorescence in situ hybridization (FISH). Springer, 2017
9. Shi L., Yang F., Kumamoto A. The chromosomes of tufted deer (Elaphodus cephalophus). *Cytogenet Genome Res*. 1991;56(3-4):189-192
10. Proskuryakova A.A. et al. Comparative studies of karyotypes in the Cervidae family. *Cytogenet Genome Res*. 2022;162(6):312-322
11. Proskuryakova A.A. et al. Comparative studies of X chromosomes in Cervidae family. *Sci Rep*. 2023;13(1):11992

doi 10.18699/bgrs2024-7.1-07

Компьютерное моделирование биотехнологических показателей аквакультуры на примере предсказания выживаемости сибирского осетра

Котов И.И.[1]*, Байрамкулов Д.Д.[1], Орлов Ю.Л.[1, 2]**

[1] *Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

** mr_lorr_40t@mail.ru; ** orlov@bionet.nsc.ru*

Ключевые слова: медицинская информатика; медицинская генетика; заболевания глаз; миопия; базы данных; генные онтологии

Мотивация и цель: Глобальные вызовы для биотехнологии связаны с возрастающей численностью населения планеты, что требует искать пути повышения производительности сельского хозяйства, улучшать сорта растений и породы животных, в том числе и рыб. Однако традиционные методы селекции дают ограниченный ответ на решение проблем, связанных с продовольственной безопасностью. По расчетам исследователей для обеспечения нормального уровня питания населения планеты в 2025 г. необходимо будет увеличить объем пищевой продукции в 2 раза [1], а вылов рыб в океанах – в 7 раз. В настоящее время аквакультура является самым быстрорастущим продовольственным сектором в мире – уже сейчас она поставляет почти половину съедобной рыбы, обеспечивая около 17 кг на душу населения. В мировой аквакультуре выращивают свыше 230 видов рыб, что намного больше по сравнению с сельскохозяйственными животными. Значительный вклад в развитие аквакультуры вносят методы современной биотехнологии. В нашей работе мы рассматриваем математические модели роста сибирского осетра при разведении в искусственных водоемах, в зависимости от структуры корма. Практическая цель данной работы – увеличить процентный состав выживаемости молодняка, используя общедоступные данные секвенирования следующего поколения вместе с метаболомными данными, для минимизации влияние фактор окружающей среды, а также внутренних болезней, на рост и развитие молодняка осетра сибирского.

Цели компьютерного исследования заключались в том, чтобы: установить ассоциации между различными модальными данными; построить модель прогнозирования признаков и оценить их точность; оценить вклад каждого информационного уровня, в прогнозирование различных признаков; получить новое представление о механизмах, лежащих в основе эффективности кормов.

Методы и алгоритмы: Определялись ключевые показатели, определяющие влияние выбора корма на рост и развитие мышечной массы. Для этого были рассмотрены 14 разнообразных слоёв мультомики и клинических ковариат, чтобы оценить эффективность использования кормов (FE) и связанные с ней эффективности производительности рыбы *Oncorhynchus tshawytscha* [2, 3].

Осетр Сибирский играет важную составляющую в создании различных продуктов питания. Однако средний возраст фертильности женской особи составляет более пяти, а то и десяти лет, в то же время выживаемость мальков составляет не более

двадцати процентов от первоначального количества. В связи с этим актуальной задачей является увеличения выживаемости и прироста молодняка, в искусственно созданных «Осетровых фермах».
Показатель кормовой эффективности (FE) – показатель, позволяющий выявить эффективность использования корма, для продуктивных потребностей организма. Потребности могут выражаться по-разному: рост, связанная с этим метаболическая активность-направленная на увеличение мышечной массы. Также показатель показывает соотношение между потреблением пищи и приростом мышечной массы, что в свою очередь является и выражается как коэффициент конверсии корма (FCR). Важно отметить инверсионность коэффициентов: низкие значения FCR, соответствуют высоким значениям FE, в то же время высокие значения FCR соответствуют низким значениям FE. Важно правильно оценивать значения FCR, поскольку на практике, в промышленных системах, зачастую уровень затрат на рыбные корма составляет 50–60 %, от общего бюджета, что в свою очередь может превышать требуемый минимум, достаточный для требуемых значения FE.
Результаты: Собрана научная литература по генам роста осетровых. Представлена математическая модель роста рыбы в зависимости от структуры питания. FE является достаточно сложным показателем, зависящим от различных связанных факторов, такие как социальное поведение, потребление корма, его переваривание а также ассимиляция питательных веществ, метаболическая активность и генетика. Выявлено что виды и высоким значениям FE, питаются маленькими порциями, тем самым эффективнее расходуя полученный белок, углеводы и липиды, что в свою очередь приводит к наилучшему конечному результату-высоким показателям мышечной массы, в отношении потребляемого корма.
Чтобы рассчитать значение FE для данных мультиомики и клинических ковариат, использовался вариант разновидности алгоритма дерева решений случайного леса (random forest – RF). RF-метод машинного обучения на основе ансамбля, способный выявлять сложные закономерности зависимости между фенотипическими признаками и ковариатами, благодаря чему широко используется для задач классификации и регрессии. Устойчивость к выбросам позволяет обрабатывать гетерогенные типы данных, без потребности к их нормализации, что в свою очередь делает его устойчивым к переобучению, а также позволяет распознавать нелинейные связи и вычислять их характеристики. Однако текущая модель RF, ограничивалась одной матрицей, при объединении данных мультиомики и не предполагается на распределения данных. Текущую проблему решает представленная модель RF[1], рандомизирую выбор слоев для каждого типа узла, а также сохраняя распределение отдельных уровней данных.
одного и того же образца (мышечная метаболомика и мышечная протеомика). Кроме-того установлена значимая корреляция между биомаркёрами крови и физическим здоровьем. Были выявлены отрицательные значения FE, при наличие мышечного глицерина, что может указывать на высокие показатели скорости использования субстратов липолиза, в свою очередь более низкие показатели FE, могут свидетельствовать о использовании глицерина в качестве субстрата. Интересно что глицерин является наиболее распространённым компонентом кормовой добавки, вытесняя аминокислоты и перенаправляя метаболический

поток в мышцах в сторону глюконеогенеза, тем самым открывая возможность использования аминокислот в полезных для организма процессах.
Выводы: В результате оценки собранных слоёв, было обнаружено значительное соответствие между многими слоями. Наибольшее соответствие было выявлено между слоями, собранными с использованием одного подхода (метаболомные плазменные показатели и мышечные), а также между слоями, полученными из одного и того же образца (мышечная метаболомика и мышечная протеомика). Кроме-того установлена значимая корреляция между биомаркерами крови и физическим здоровьем. Были выявлены отрицательные значения FE, при наличие мышечного глицерина, что может указывать на высокие показатели скорости использования субстратов липолиза, в свою очередь более низкие показатели FE, могут свидетельствовать о использовании глицерина в качестве субстрата. Интересно что глицерин является наиболее распространённым компонентом кормовой добавки, вытесняя аминокислоты и перенаправляя метаболический поток в мышцах в сторону глюконеогенеза, тем самым открывая возможность использования аминокислот в полезных для организма процессах.
Вывод текущего исследования строиться на важности индивидуального измерения показателей FE, что в свою очередь позволяет персонализировано оптимизировать затраты под выбранного рыбного представителя. Это позволит грамотно распределять выделенные ресурсы бюджета на кормовые добавки, тем самым контролируя производственных выход здорового и эффективного метаболистические поколения молодых осетров. Данных подход является достаточно дорогостоящим, поскольку требует персонализированного исследования, анализа омики, в том числе и метаболомики, а также проведение секвенирования следующего поколения (NGS).
Благодарности: Авторы благодарны сотрудникам аквафермы «Осетровые фермы», Романовское хозяйство, Калужская область, за организацию работы.

Computer modeling of biotechnological indicators of aquaculture on the example of predicting the survival rate of Siberian sturgeon

Kotov I.I.[1]*, Bayramkulov D.D.[1], Orlov Y.L.[1, 2]**

[1] *Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia*

[2] *Institute of Cytology and Genetics SB RAS Novosibirsk, Russia*

** mr_lorr_40t@mail.ru, ** orlov@bionet.nsc.ru*

Key words: medical informatics*;* medical genetics*;* eye diseases*;* myopia*;* databases*;* gene ontologies

Motivation and Purpose: Global challenges for biotechnology are related to the increasing world population, which requires finding ways to increase agricultural productivity, improve plant varieties and animal breeds, including fish. However, traditional breeding methods provide a limited answer to solving problems related to food security. According to researchers' calculations, to ensure a normal level of nutrition of the world population in 2025, it will be necessary to increase the volume of food production by 2 times [1], and fish catch in the oceans – by 7 times. Aquaculture is currently the fastest growing food sector in the world – already supplying nearly half of

the world's edible fish, providing about 17 kg per capita. More than 230 species of fish are farmed in global aquaculture, a far greater number than farm animals. The methods of modern biotechnology make a significant contribution to the development of aquaculture. In our work we consider mathematical models of growth of Siberian sturgeon during breeding in artificial reservoirs, depending on the feed structure. The practical goal of this work is to increase the percentage of juvenile survival using publicly available next generation sequencing data together with metabolomic data to minimize the influence of environmental factors, as well as internal diseases, on the growth and development of juvenile Siberian sturgeon.

The objectives of the computerized study were to: establish associations between different modal data; build a trait prediction model and assess its accuracy; evaluate the contribution of each information level, to the prediction of different traits; and gain new insights into the mechanisms underlying feed efficiency.

Methods and Algorithms: Key indicators determining the effect of feed selection on muscle mass growth and development were identified. For this purpose, 14 diverse layers of multomics and clinical covariates were considered to assess feed utilization efficiency (FE) and associated performance efficiencies of fish Oncorhynchus tshawytscha [2, 3].

Siberian sturgeon plays an important component in the development of various food products. However, the average age of fertility of female fish is more than five or even ten years, at the same time the survival rate of fry is not more than twenty percent of the original number. In this regard, the urgent task is to increase the survival rate and growth of young fish in artificially created “Sturgeon farms”

Feed efficiency (FE) is an indicator that reveals the efficiency of feed utilization for the productive needs of the organism. The needs can be expressed in different ways: growth, associated metabolic activity-directed to increase muscle mass. It also shows the relationship between food intake and muscle mass gain, which in turn is and is expressed as feed conversion ratio (FCR). It is important to note the inversion of the ratios: low FCR values correspond to high FE values, while high FCR values correspond to low FE values. It is important to correctly estimate FCR values, because in practice, in industrial systems, the level of fish feed costs is often 50–60 % of the total budget, which in turn may exceed the required minimum, sufficient for the required FE values.

Results: Scientific literature on sturgeon growth genes was collected. A mathematical model of fish growth in relation to feeding pattern is presented. FE is a rather complex index depending on various related factors such as social behavior, feed intake, feed digestion as well as nutrient assimilation, metabolic activity and genetics. Species with high FE values are found to eat smaller portions, thus more efficiently utilizing the protein, carbohydrates and lipids obtained, which in turn leads to the best end result-high muscle mass in relation to feed intake

To calculate the FE value for multivariate data and clinical covariates, a variant of the random forest (RF) decision tree algorithm was used.

An ensemble-based RF machine learning method capable of detecting complex patterns of dependence between phenotypic traits and covariates, making it widely used for classification and regression tasks. Its robustness to outliers allows it to handle heterogeneous data types without the need for normalization, which in turn makes it robust to overfitting, and also allows it to recognize nonlinear relationships and compute their characteristics. However, the current RF model, was limited to a single matrix, when combining multi-omics data and is not assumed on data distributions. The current

problem is solved by the presented RF model [1], by randomizing the choice of layers for each node type, and preserving the distribution of individual data layers.
of the same sample (muscle metabolomics and muscle proteomics).
In addition, there was a significant correlation between blood biomarkers and physical health. Negative FE values were found in the presence of muscle glycerol, which may indicate high rates of lipolysis substrate utilization, while lower FE values may indicate the use of glycerol as a substrate. Interestingly, glycerol is the most common component of feed additives, displacing amino acids and redirecting the metabolic flux in muscle toward gluconeogenesis, thus opening up the possibility of utilizing amino acids in beneficial processes.
Conclusions: From the evaluation of the collected layers, significant concordance was found between many layers. The highest concordance was found between layers collected using the same approach (plasma metabolomic indices and muscle metabolomic indices) and between layers derived from the same sample (muscle metabolomics and muscle proteomics). In addition, a significant correlation was found between blood biomarkers and physical health. Negative FE values were found in the presence of muscle glycerol, which may indicate high rates of lipolysis substrate utilization, while lower FE values may indicate the use of glycerol as a substrate. Interestingly, glycerol is the most common component of feed additives, displacing amino acids and redirecting the metabolic flux in muscle towards gluconeogenesis, thus opening up the possibility of utilizing amino acids in beneficial processes. The conclusion of the current study is based on the importance of individual measurement of FE values, which in turn allows personalized cost optimization for the selected fish representative. This will allow a competent allocation of allocated budget resources for feed additives, thus controlling the production output of healthy and efficient metabolic generation of young sturgeons. This approach is quite expensive as it requires personalized research, omics analysis, including metabolomics, and next generation sequencing (NGS).
Acknowledgments: The authors are grateful to the staff of Sturgeon Farms Aquafarm, Romanovskoye farm. Tver region, for organizing the work.

Список литературы/References

1. Young T., Laroche O., Walker S.P. et al. Prediction of feed efficiency and performance-based traits in fish via integration of multiple omics and clinical covariates. *Biology*. 2023;12(8):1135. doi 10.3390/biology12081135
2. Hung S.S.O. Recent advances in sturgeon nutrition. *Anim Nutr*. 2017;3:191-204. doi 10.1016/j.aninu.2017.05.005
3. Zhu Y. et al. Metabolomics and gene expressions revealed the metabolic changes of lipid and amino acids and the related energetic mechanism in response to ovary development of chinese sturgeon (Acipenser Sinensis). *PLoS One*. 2020;15:e0235043. doi 10.1371/journal.pone.0235043

doi 10.18699/bgrs2024-7.1-08

Изменение транскриптома гиппокампа неонатальных крыс после введения синтетического глюкокортикоида дексаметазона

Ланшаков Д.[1, 2]*, Сборщикова А.[1, 2], Сухарева Е.[1], Булыгина В.[1], Калинина Т.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *lanshakov@bionet.nsc.ru*

Ключевые слова: глюкокортикоиды; гиппокамп; неонатальный период; транскриптом

Мотивация и цель: Глюкокортикоиды – гормоны стресса – являются одними из важнейших факторов, влияющих в перинатальный период на развитие головного мозга млекопитающих. В перинатальной медицине препараты глюкокортикоидов используются для предотвращения осложнений, связанных с преждевременными родами, для ускорения выработки сурфактанта легкими. Такая терапия, назначаемая по жизненным показаниям, имеет долговременные осложнения, связанные как с общим физиологическим развитием, так и развитием ЦНС, когнитивных способностей и памяти. Целью исследования было выявить возможные гены-мишени, вовлеченные в этот процесс.

Методы и алгоритмы: Для этого нами были исследованы изменения транскриптома гиппокампа неонатальных крыс через 6 ч после введения синтетического глюкокортикоида дексаметазона в дозе 0.2 мг/кг веса тела на второй день жизни P2, день родов считали нулевым днем P0. Контрольная группа получала инъекцию такого же объема физиологического раствора. Методом массового параллельного секвенирования транскриптома (100M парных прочтений на платформе Illumina) было выявлено изменение экспрессии основных генов-мишеней. Технические последовательности обрезались при помощи программы Триммоматик. Риды выравнивались на референсный геном крысы mRatBN7.2 при помощи картировщика hisat2. Для подсчета ридов, приходящихся на кодирующую часть генов, использовали программу Rsubread. Дифференциально экспрессированные гены определяли при помощи пакета edger. Функциональную аннотация проводили при помощи пакета clusterprofiler.

Результаты: Так, выбрав дифференциально экспрессированные гены $|logFC|>1$ $p.adj < 0.05$ (с поправкой на множественность сравнений Бенджамини–Хохберга), изменили экспрессию 57 генов: 46 увеличили, 11 снизили экспрессию (рис. 1, A). Гены, чья экспрессия увеличилась, принадлежали к таким разделам «Биологические процессы» классификации Генной Онтологии (GO): клеточный ответ на глюкокортикоидный стимул, клеточный ответ на кортикостероидный стимул – что естественно при интенсификации глюкокортикоидного сигналинга (см. рис. 1, B). Также были выявлены следующие категории: эйкозаноид метаболические процессы, детоксификация, стрессовый ответ на ионы металлов, клеточный ответ на ионы цинка (см. рис. 1, C). Гены, снизившие экспрессию, показали обогащение среди следующих категорий Генной Онтологии: клеточный ответ на липополисахарид, клеточный ответ на молекулы бактериальной природы,

клеточный ответ на биотические стимулы, позитивная регуляция пролиферации NK-клеток в ответ на LPS, дегрануляция NK-клеток.

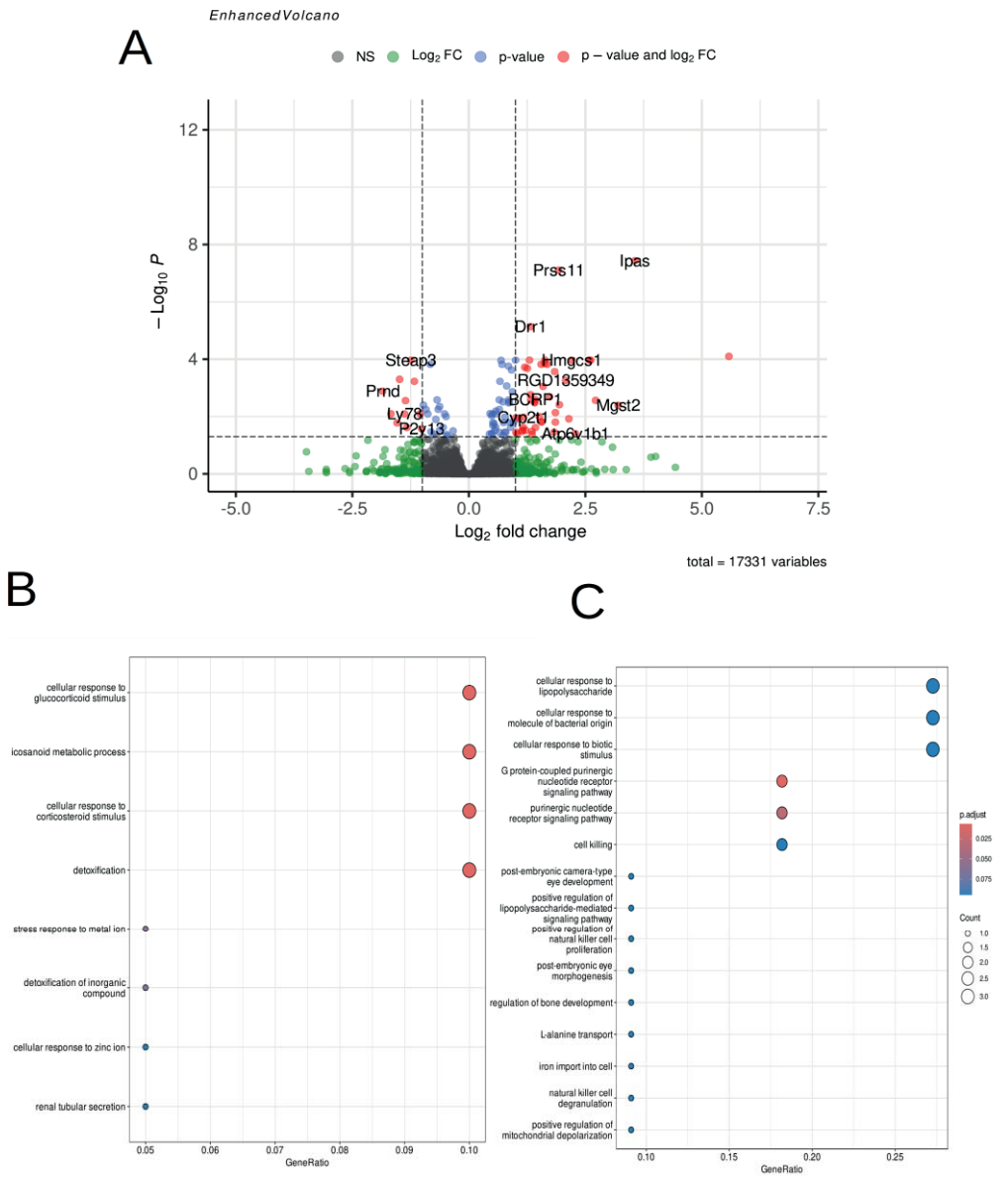


Рис. 1. Характеризация транскриптомных изменений в гиппокампе двухдневных крыс через 6 ч после введения дексаметазона 0.2 мг/кг подкожно. A – вулкано плот, показывающий дифференциально экспрессирующиеся гены, красным отмечены гены интереса |log2FC| > 1 & p.adj(BH) < 0.05; B – функциональная аннотация и обогащение по GO «Биологические процессы» генов, увеличивших экспрессию; C – функциональная аннотация и обогащение по GO «Биологические процессы» генов, снизивших экспрессию

Данные категории в первую очередь связаны с действием глюкокортикоидов на иммунные клетки и резидентные макрофаги мозга клетки микроглии. Стоит отметить, что среди GO категорий также были: сигнальный путь пуринэргических нуклеотидных рецепторов, связанных с G-белками, транспорт L-аланина, импорт железа в клетку, позитивная регуляция деполяризации митохондрий. Среди генов, которые повысили свою экспрессию, были Фактор, Индуцируемый Гипоксией, 3A (Hif3a) logFC = 3.56, который может регулировать плюрипотентность и пролиферацию эмбриональных стволовых клеток [1]. Также экспрессия серотонинового рецептора 1A (Htr1a) была повышена logFC = 1.93. Это может быть компенсаторным эффектом снижения экспрессии Tph2 в среднем мозге и серотонина в передних отделах после введения дексаметазона [2]. Ранее нами было показано, что однократная инъекция на второй день жизни крыс приводит к уменьшению тревожного и депрессивно-подобного поведения на 60-й день жизни [3]. Это может объясняться одновременной интенсификацией норадренэргической нейротрансмиссии [4] и снижением серотонинэргической нейротрансмиссии [2], что приводит к данному поведенческому фенотипу. Схожие эффекты на поведение наблюдаются при снижении серотониновой нейротрансмиссии в неонатальный период [5]. Несмотря на выраженную супрессию иммунной функции, по результатам обогащения ГО, после введения дексаметазона, были повышены Интерлейкин-33 IL-33 logFC = 1.59 и Маннозный рецептор C-типа 1 (Mrc1) logFC = 1.30. Эти гены являются маркерами M2 микроглии, которая запускает тканевую реорганизацию и регенерацию, увеличивает фагоцитоз и играет нейропротективную роль [6]. Среди генов, снизивших свою экспрессию, стоит отметить второй рецептор васкулярного эндотелиального ростового фактора Kdr logFC = –1.04, который ответственен за рост аксонов в ответ на Семафорин 3E (Sema3E) [7].

Выводы: Таким образом, помимо выявленных классических эффектов дексаметазона на экспрессию генов иммунной системы, возможными генами-мишенями для изучения негативных эффектов глюкокортикоидов на развитие нервной системы являются Hif3a, Htr1a, Kdr. Показанное нами ранее снижение тревожного и депрессивно-подобного поведения на 60-й день жизни крыс после однократной инъекции дексаметазона на P2 может объясняться одновременной интенсификацией норадренэргической нейротрансмиссии и снижением серотонинэргической нейротрансмиссии.

Финансирование: Исследование поддержано бюджетным проектом FWNR-2023-0002.

Alteration of the hippocampal transcriptome in neonatal rats after administration of the synthetic glucocorticoid dexamethasone

Lanshakov D.[1, 2]*, Sborshchikova A.[1, 2], Sukhareva E.[1], Bulygina V.[1], Kalinina T.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *lanshakov@bionet.nsc.ru*

Key words: glucocorticoids; hippocampus; neonatal period; transcriptome

Motivation and Aim: Glucocorticoids, stress hormones, are one of the most important factors affecting the development of the mammalian brain during the perinatal period. In perinatal medicine, Glucocorticoid medications are used to prevent complications associated with premature birth, to accelerate the production of surfactant by the lungs. Such therapy, prescribed based on vital signs, has long-term complications associated with both general physiological development and the development of the central nervous system, cognitive abilities and memory. The aim of the study was to identify possible target genes involved in this process.
Methods and Algorithms: To do this, we investigated transcriptomic changes in the hippocampal of neonatal rats 6 hours after administration of the synthetic glucocorticoid dexamethasone at a dose of 0.2 mg/kg body weight on the second day of life P2, the day of delivery was considered the zero day P0. The control group received an injection of the same volume of saline solution. The method of mass parallel sequencing of the transcriptome (100M paired reads on the Illumina platform) revealed a change in the expression of the main target genes. The technical sequences were cut using the Trimmomatic program. The reads were aligned to the reference rat genome mRatBN7.2 using the hisat2 aligner. The Rsubread program was used to calculate the reads aligned to the coding part of genes. Differentially expressed genes were determined using the edge R package. The functional annotation was carried out using the clusterprofiler R package.
Results: Thus, by selecting differentially expressed genes $|logFC|>1$ p.adj < 0.05 (adjusted for the Benjamin-Hochberg multiple comparison corection), 57 genes were changed in expression; 46 increased, 11 decreased expression (Fig. 1A). Genes whose expression increased belonged to such terms of the "Biological processes" classification of Gene Ontology (GO): cellular response to a glucocorticoid stimulus, cellular response to a corticosteroid stimulus – which is natural with the intensification of glucocorticoid signaling (Fig. 1B). The following categories were also identified: icosanoid metabolic processes, detoxification, stress response to metal ions, cellular response to zinc ions (Fig. 1C). Genes that reduced expression showed enrichment among the following categories of Gene Ontology: cellular response to lipopolysaccharide, cellular response to bacterial molecules, cellular response to biotic stimuli, positive regulation of NK cell proliferation in response to LPS, degranulation of NK cells. These categories are primarily related to the effect of glucocorticoids on immune cells and brain resident macrophages of microglial cells. It is worth noting that among the GO categories were also: the signaling pathway of purinergic nucleotide receptors associated with G proteins, L-alanine transport, iron import into the cell, positive regulation of mitochondrial depolarization.
It is worth noting that among the genes that increased their expression were Hypoxia-induced Factor 3A (Hif3a) logFC = 3.56, which can regulate pluripotency and proliferation of embryonic stem cells [1]. Also, the expression of serotonin receptor 1A (Htr1a) was increased logFC = 1.93. This may be a compensatory effect of a decrease in the expression of Tph2 in the midbrain and serotonin level in the anterior regions after administration of dexamethasone [2]. Previously, we showed that a single injection on the second day of rat life leads to a decrease in anxiety and depressive-like behavior on the 60th day of life [3]. This can be explained by the simultaneous intensification of noradrenergic neurotransmission [4] and a decrease in serotonergic neurotransmission

[2], which leads to this behavioral phenotype. Similar effects on behavior are observed with a decrease in serotonin neurotransmission in the neonatal period [5].

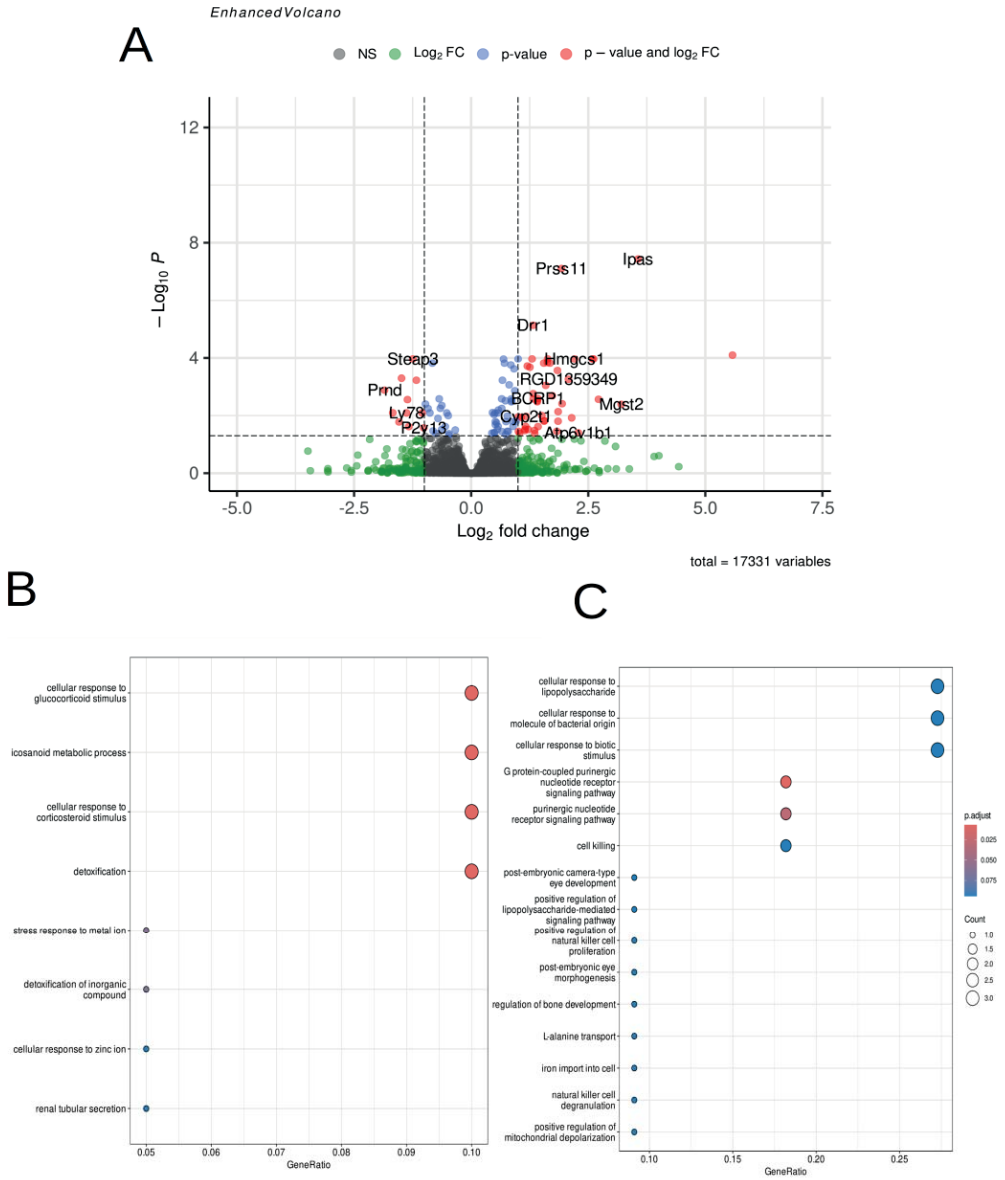


Fig. 1. Characterization of transcriptomic changes in the hippocampus of two-day-old rats 6 hours after administration of dexamethasone 0.2 mg/kg subcutaneously. A – volcano plot showing differentially expressed genes, the genes of interest are marked in red |log2FC| > 1 & p.adj(BH) < 0.05; B – functional annotation and GO terms "Biological processes" enrichment of genes that have increased expression; C – functional annotation and GO terms "Biological processes" enrichment of genes that have decreased expression

Despite the pronounced suppression of immune function, according to the results of GO enrichment, after administration of dexamethasone, Interleukin-33 IL-33 logFC = 1.59 and Mannose receptor C-type 1 (Mrc1) logFC = 1.30 were increased. These genes are markers of M2 microglia, which triggers tissue reorganization and regeneration, increases phagocytosis and plays a role neuroprotective role [6]. Among the genes that have decreased their expression, it is worth noting the second receptor of the vascular endothelial growth factor Kdr logFC = –1.04, which is responsible for the growth of axons in response to Semaphorin 3E (Sema3E) [7].

Conclusion: Thus, in addition to the identified classical effects of dexamethasone on the expression of immune system genes, possible target genes for studying the negative effects of glucocorticoids on the development of the nervous system are Hif3a, Htr1a, Kdr. The previously shown by us decrease in anxiety and depressive-like behavior on the 60th day of rat life after a single injection of dexamethasone on P2 may be explained by the simultaneous intensification of noradrenergic neurotransmission and a decrease in serotonergic neurotransmission.

Funding: The study is supported by Ministry of Education project FWNR-2023-0002.

Список литературы/References

1. Forristal C.E. et al. Hypoxia inducible factors regulate pluripotency and proliferation in human embryonic stem cells cultured at reduced oxygen tensions. *Reproduction.* 2010;139:85-97. doi 10.1530/REP-09-0300
2. Clark J.A. et al. Glucocorticoid modulation of tryptophan hydroxylase-2 protein in raphe nuclei and 5-hydroxytryptophan concentrations in frontal cortex of C57/Bl6 mice. *Mol Psychiatry.* 2008;13:498-506. doi 10.1038/sj.mp.4002041
3. Lanshakov D.A. et al. Single neonatal dexamethasone administration has long-lasting outcome on depressive-like behaviour, Bdnf, Nt-3, P75ngfr and sorting receptors (SorCS1-3) stress reactive expression. *Sci Rep.* 2021;11:8092. doi 10.1038/s41598-021-87652-7
4. Kalinina T.S., Shishkina G.T., Dygalo N.N. Induction of tyrosine hydroxylase gene expression by glucocorticoids in the perinatal rat brain is age-dependent. *Neurochem Res*. 2012;37:811-818. doi 10.1007/s11064-011-0676-y
5. Trujillo V. et al. Neonatal serotonin depletion induces hyperactivity and anxiolytic-like sex-dependent effects in adult rats. *Mol Neurobiol.* 2021;58:1036-1051. doi 10.1007/s12035-020-02181-0
6. Cherry J.D., Olschowka J.A., O'Banion M.K. Neuroinflammation and M2 microglia: the good, the bad, and the inflamed. *J Neuroinflammation.* 2014;11:98. doi 10.1186/1742-2094-11-98
7. Bellon A. et al. VEGFR2 (KDR/Flk1) Signaling mediates axon growth in response to semaphorin 3E in the developing brain. *Neuron.* 2010;66:205-219. doi 10.1016/j.neuron.2010.04.006

doi 10.18699/bgrs2024-7.1-09

Метилсульфонилметан (MCM) вызывает повышение экспрессионного (генетического) шума, но не влияет на плодовитость и продолжительности жизни *D. melanogaster*

Орбант М.О.[1], Бабочкина Т.И.[1], Фёдорова С.А.[1, 2]*

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

* *fsveta@bionet.nsc.ru*

Ключевые слова: генетический шум; экспрессионный шум; MCM; $ДМСО_2$; фенотип

Мотивация и цель: Развитие любого организма обусловлено скоординированной пространственно-временной экспрессией генов, в результате которой формируется генетически предопределенный фенотип [1, 2]. Однако все этапы реализации генетической информации (транскрипция, трансляция, посттрансляционные модификации, транспорт молекул) подвержены стохастическим флуктуациям. Такие случайные колебания, возникающие при реализации генетической программы, называются генетическим (экспрессионным) шумом и могут быть важным источником вариаций на морфологическом уровне, что проявляется в вариациях фенотипов (так называемый шум развития).

Экспрессионный шум играет важную роль во многих биологических процессах: например, он повышает чувствительность эмбриональных стволовых клеток мыши к сигналам дифференцировки/дедифференцировки [3], с ним связывают устойчивость клеток к химиотерапии [4]. Таким образом, изучение влияния экспрессионного шума на реализацию наследственных признаков и способов манипулирования его вариабельностью является актуальной задачей.

Целью данной работы было оценка влияния экспрессионного шума на шум развития на модели *Drosophila melanogaster*. Для этого мы исследовали влияние разных концентраций MCM на экспрессию некоторых генов оогенеза, плодовитость и продолжительность жизни *D. melanogaster*.

Методы и алгоритмы: В работе использовали БАД метилсульфонилметан (MCM, $ДМСО_2$) – нетоксичное вещество, которое часто применяется для снятия воспалительных процессов в суставах и связках. MCM является метаболитом ДМСО, влияющим на экспрессию генов и на температуру плавления ДНК [5, 6]. В работе использовали MCM в концентрациях в корме 0, 0.01, 0.1 и 1 мг/мл.

Экспрессию генов оценивали при помощи количественного ПЦР в реальном времени (экспрессия на уровне РНК). Количество белка оценивали по интенсивности свечения химерного белка pABP2-GFP в ядрах питающих клеток и ооцитов *D. melanogaster*, для этого использовали линию мух с белковой ловушкой – искусственным экзоном, кодирующим EGFP, встроенным в ген *pAbp2*. Получившийся химерный белок экспрессируется под нативным промотором и позволяет визуализировать белковый продукт гена в реальном времени.

В качестве фенотипических признаков оценивали плодовитость самок, выживаемость мух от эмбриона до стадии имаго и продолжительности жизни взрослых мух.

Для всех экспериментов проведено не менее двух биологических повторов.

Результаты: Для оценки экспрессии генов на уровне мРНК выделяли РНК из 50 яичников мух, содержащихся на корме без МСМ, на корме с концентрациями МСМ 0.01, 0.1 и 1 мг/мл. При помощи qPCR исследовали экспрессию двух разных групп генов: повсеместно экспрессирующиеся гены (*myc*, *sep1*, *hyd*, *stat-92*) и тканеспецифичные гены (*pKa*, *Cup*, *hop*, *pABP2*), нормированных на экспрессию гена *rpl32*. И повсеместно экспрессирующиеся, и яичник-специфичные гены демонстрировали разнонаправленное изменение уровня экспрессии в разных биологических повторах (проведено два биологических повтора), не зависящее от дозы МСМ.

Для оценки экспрессии генов на уровне белка оценивали уровень светимости химерного белка при помощи линии мух с белковой ловушкой. Измеряли разницу в интенсивности свечения ядер соседних питающих клеток или ооцитов на одномоментно сфотографированном препарате, что позволило нам снизить технические проблемы нормировки. Для анализа выбирали только ядра клеток зародышевой линии из овариол 6–7-й стадии, потому что к этой стадии они заканчивают эндоцикл и имеют одинаковую плоидность. Дозозависимой разницы в интенсивности свечения химерного белка в ядрах клеток зародышевой линии не выявлено.

Для оценки плодовитости самок содержали в индивидуальных пробирках на корме с разными концентрациями МСМ и ежедневно подсчитывали количество яиц, отложенных каждой самкой. Достоверных различий в плодовитости самок, содержащихся на разных концентрациях МСМ в корме, не обнаружено.

Выживаемость до стадии имаго анализировали, подсчитывая количество эмбрионов в пробирке, затем – количество куколок и вылупившихся имаго в этой же пробирке. Выживаемость подсчитывали как % эмбрионов, развившихся до стадии взрослых мух. Продолжительность жизни измеряли, содержа по 10 самцов или самок в пробирках с различными концентрациями МСМ в корме.

Мух перекидывали на свежий корм 1 раз в 4 дня, а 1 раз в 2 дня фиксировали количество погибших мух, выстраивая кривые дожития. Проведено два биологических повтора, по 100–120 мух для каждой концентрации МСМ в каждом повторе. В диапазоне концентраций МСМ в корме 0.01–1 мг/мл дозозависимого влияния на выживаемость и продолжительность жизни не выявлено.

Выводы: Мы показали, что МСМ, как и его предшественник ДМСО, снижает температуру плавления ДНК. В связи с этим мы ожидали, что добавление МСМ в корм приведет к повышению экспрессионного шума и, возможно, к увеличению шума развития. Однако наше исследование не выявило дозозависимого влияния МСМ на экспрессию генов как на уровне мРНК, так и на уровне белка. Более того, хотя мы и фиксировали разнонаправленные колебания экспрессии генов, достоверных различий на уровне фенотипа (плодовитость самок, выживаемость эмбрионов до стадии имаго, продолжительность жизни взрослых мух) не выявлено.

Финансирование: Исследование поддержано бюджетным проектом FWNR-2022-0015.

Methylsulfonylmethane (MSM) increases the expression (genetic) noise, but has no effect on the fertility and life span of *D. melanogaster*

Orbant M.O.[1], Babochkina T.I.[1], Fedorova S.A.[1, 2]*

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *fsveta@bionet.nsc.ru*

Key words: genetic noise; expression noise; MSM; $DMSO_2$; phenotype

Motivation and Aim: The development of any organism is driven by the coordinated spatiotemporal expression of genes. The result is a genetically predetermined phenotype is formed [1, 2]. However, stochastic fluctuations occur at all stages of the realisation of genetic information (transcription, translation, post-translational modifications, transport of molecules). These random fluctuations are referred to as genetic (expression) noise, and can be an important source of variation at the morphological level, which manifests itself in variations in phenotypes (the so-called developmental noise).

Expression noise plays an important role in many biological processes: for example, it increases the sensitivity of mouse embryonic stem cells to differentiation/ dedifferentiation signals [3] and is associated with cell resistance to chemotherapy [4]. Therefore the study of the influence of expression noise on the realisation of heritable traits and the development of the ways to manipulate its variability is an urgent current task.

The aim of this work was to assess the influence of expression noise on developmental noise in the *Drosophila melanogaster* model. To do this, we studied the effect of different concentrations of MSM on the gene expression (oogenesis genes), the fertility and the lifespan of *D. melanogaster*.

Methods and Algorithms: We used the dietary supplement methylsulfonylmethane (MSM, $DMSO_2$), which is a non-toxic substance and is commonly used in the treatment of rheumatoid diseases to alleviate inflammatory processes in joints and ligaments. MSM ($DMSO_2$) is a metabolite of DMSO that affects gene expression and DNA melting temperature [5, 6]. In our study, we used dietary MSM at concentrations of 0 mg/ml, 0.01 mg/ml, 0.1 mg/ml, and 1 mg/ml.

Gene expression was assessed by quantitative real-time PCR (RNA level expression). The amount of protein was assessed by the fluorescence intensity of the chimeric protein pABP2-GFP in the nuclei of *D. melanogaster* nurse cells and oocytes; using a fly strain with a protein trap – an artificial exon encoding EGFP, inserted into the second intron of the *pAbp2* gene. The resulting chimeric protein is expressed under the native promoter, allowing real-time visualization of the gene product.

As phenotypic traits, we assessed female fertility, fly survival from the embryo to the imago stage, and the lifespan of adult flies.

At least 2 biological replicates were performed for each experiment.

Results: In order to assess gene expression at the mRNA level, RNA was isolated from 50 ovaries of flies from the following groups: maintained on the diet without MSM, on the diet with MSM at concentrations of 0.01, 0.1 and 1 mg/ml. Using qPCR, we studied

the expression of two different groups of genes: ubiquitously expressed genes (*myc*, *sep1*, *hyd*, *stat-92*) and tissue-specific genes (*pKa*, *Cup*, *hop*, *pABP2*), normalized to the expression of the rpl32 gene. Both ubiquitously expressed and ovary-specific genes showed multidirectional changes in expression levels in different biological replicates (2 biological replicates were performed), independent of MSM dose.

To assess gene expression at the protein level, the fluorescence intensity level of the chimeric protein was assessed using a protein trap fly line. We measured the difference in the fluorescence intensity of the nuclei from neighboring nurse cell or oocyte nuclei on a simultaneously photographed preparation, which allowed us to reduce the technical problems of normalization. Only germ cell nuclei from stage 6–7 ovarioles were selected for analysis, as they have completed the endocycle at this stage and have the same ploidy. No dose-dependent difference in the fluorescence intensity of the chimeric protein in the nuclei of germline cells was observed.

To assess fecundity, females were kept in individual tubes with different concentrations of MSM in the diet and the number of eggs laid by each female was counted daily. No significant differences were found in the fertility of females kept on different concentrations of MSM in the diet.

Survival to the adult stage was analysed by counting the number of embryos in a test tube, then the number of pupae and hatched adults in the same test tube. Survival was calculated as the percentage of embryos that developed into adult flies. Longevity was measured by keeping 10 males or females in test tubes containing different concentrations of MSM in the diet.

Flies were transferred to fresh food every 4 days and the number of dead flies was recorded every 2 days to construct survival curves. Two biological replicates were performed, with 100–120 flies for each MSM concentration in each replicate. There was no dose-dependent effect on survival and longevity over the range of MSM concentrations in the diet from 0.01–1 mg/ml.

Conclusion: We have shown that MSM, like its precursor DMSO, reduces DNA melting temperature. In this context, we expected that the addition of MSM to the diet would lead to an increase in expression noise and possibly an increase in developmental noise. However, our study did not show a dose-dependent effect of MSM on gene expression at both mRNA and protein levels. Furthermore, although we observed multidirectional fluctuations in gene expression, no significant differences were found at the phenotypic level (fecundity of females, survival of embryos to the adult stage, lifespan of adult flies).

Funding: The study was supported by the budget project FWNR-2022-0015.

Список литературы/References

1. Waddington C. Canalization of development and the inheritance of acquired characters. *Nature*. 1942;150:563-565. doi 10.1038/150563a0
2. Dong P., Liu Z. Shaping development by stochasticity and dynamics in gene regulation. *Open Biol*. 2017;7:170030. doi 10.1098/rsob.170030
3. Desai R.V. et al. A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. *Science*. 2021;373(6557):eabc6506. doi 10.1126/science.abc6506
4. Jordan F.H. et al. Memory of stochastic single-cell apoptotic signalling promotes chemoresistance in neuroblastoma. *Sci Adv*. 2023;9(9):eabp8314. doi 10.1126/sciadv.abp8314
5. Stephens K.E. et al. Epigenetic regulation and measurement of epigenetic changes. *Biol Res Nurs*. 2013;15(4):373-381. doi 10.1177/1099800412444785
6. Butawan M. et al. Methylsulfonylmethane: Applications and Safety of a Novel Dietary Supplement. *Nutrients*. 2017;9(3):290. doi 10.3390/nu9030290

doi 10.18699/bgrs2024-7.1-10

Эволюция тандемных повторов на примере двух групп позвоночных животных

Остромышенский Д.И.[1]*, Попова М.А.[1, 2], Травина А.О.[1]

[1] *Институт цитологии РАН, Санкт-Петербург, Россия*

[2] *Центр молекулярной и клеточной биологии, Сколковский институт науки и технологий, Москва, Россия*

* *necroforus@gmail.com*

Ключевые слова: тандемные повторы; гетерохроматин; хомяки; лягушки; эволюция

Мотивация и цель: Тандемные повторы (ТП) являются важной частью генома всех позвоночных животных. Конститутивный гетерохроматин, формирующий такие функционально значимые части хромосом, как центромерные, перицентромерные и субтеломерные районы, состоит в основном из ТП. ТП состоят из многократно повторящихся мономеров длиной от нескольких п.н. до нескольких т.п.н. Эти мономеры формируют поля ТП длиной до нескольких десятков м.п.н. Из-за особенностей организации и сложности сборки ТП зачастую недопредставлены в геномных сборках. Эволюция ТП также является крайне малоизученным вопросом. Появление большого количества сборок геномов, сделанных с использованием технологий секвенирования с длинными ридами (PacBio и OxfordNanopore), позволяет более эффективно исследовать ТП и их эволюцию.
Методы и алгоритмы: Искали ТП в сборках геномов лягушки *Rana temporaria* (WGS project CAJIMO01), хомяков *Cricetulus griseus* (RAZU01), *Mesocricetus auratus* (JAFVMI01), *Phodopus sungorus* (JANBXA01) и *P. roborovskii* (CALSGD01) по ранее описанной методике [1]. Рассматривали ТП с длиной поля >10 т.п.н., кроме представителей *Phodopus* у которых рассматривали поля ТП с длиной >5 т.п.н. Найденные мономеры семейств ТП искали в доступных геномах всех видов семейства Ranidae для лягушки (*R. temporaria*, *R. kukuronis*, *R. muscosa*, *Lythobathes sylvaticus*, *L. septentrionalis*, *Aquarana catesbeiana*, *Staurois parvus* и *Glandirana rugosa*) и подсемейства Cricetinae для хомяков (указанные выше виды). Поиск осуществляли с помощью BLAST с параметрами “dust” “no”, “evalue” “1-e4”. Из дальнейшего анализа удаляли хиты BLAST с bit-score < 50 и соотношением bit-score/длина мономера ТП < 1.
Результаты: В геноме *R. temporaria* найдено 77 семейств ТП. Среди них не нашли ярко выраженного мажорного по содержанию ТП. 6 семейств ТП примерно одинаковы по суммарной длине полей. Только 15 из найденных семейств ТП являются видоспецифичными. В том числе три из числа наиболее представленных в геноме Rtem1149A, Rtem138A и Rtem35A. Еще 14 семейств ТП найдены также у близкородственного вида *R. kukuronis*, в том числе один из мажорных для *R. temporaria* семейств Rtem494A. Одно из этих семейств – Rtem149A в геноме *R. kukuronis* присутствует в единичных копиях в отличие от генома *R. temporaria*. В геномах всех трех представителей Rana, но не других родов найдено 2 семейства ТП Rtem103A и Rtem965A, но у *R. kukuronis* и *R. muscosa* эти семейства представлены единичными копиями. Еще один из мажорных ТП *R. temporaria* Rtem27A представлен также большим числом копий (более 10 тысяч) и у других

Rana, но единичными копиями у видов других родов Ranidae. Около 10 семейств ТП *R. temporaria* имеют сходное содержание в исследованных геномах *Ranidae*, что может говорить о том, что эти ТП произошли от диспергированных повторов. У хомяков (Cricetinae) нашли от 45 (*P. roborovskii*) до 65 (*P. sungorus*) семейств ТП. У всех хомяков обнаружены ярко выраженные мажорные семейства ТП Cgri33A (*C. griseus*), Maur49A (*M. auratus*) или несколько мажорных ТП – Prob863A, Prob18A и Prob19A (*P. roborovskii*); Psun20A, Psun1212A, Psun32A (*P. sungorus*). У всех исследованных видов около половины семейств ТП являются видоспецифичными и не найдены у других видов. В большинстве случаев к числу невидоспецифичных относятся семейства ТП с относительно небольшим содержанием в геноме. Исключением является один из мажорных ТП *P. roborovskii* Prob18A, который найден в геномах трех других видов, но в количестве не более 100 копий, что значительно меньше копийности в геноме *P. roborovskii.* Другие семейства ТП с большим содержанием в геномах либо не встречаются у других видов, либо представлены небольшим количеством копий, к числу таких семейств относятся, например, ТП *C. griseus* Cgri72A, Cgri23A; ТП *M. auratus* Maur24A, Maur20A и др. Для рода *Phodopus* нашли 5 родоспецифичных семейств ТП. Также у видов этого рода нашли значительное количество видоспецифичных ТП с мономером длиной около 1 т.п.н. и более, образующих поля длиной более 100 т.п.н. ТП с такими длинными мономерами часто являются производными диспергированных повторов [1, 2], но найденные нами ТП не показали гомологии с диспергированными повторами.

Выводы: Наиболее популярной моделью эволюции ТП является библиотечная гипотеза [3]. Появление сборок геномов близкородственных видов позволяет проверить эту гипотезу на полном наборе ТП видов разной степени филогенетической близости. Показано, что паттерн распределения ТП в геномах Ranidae в целом соответствует библиотечной гипотезе, в то время как у хомяков противоречит ей. Это может свидетельствовать о том, что эволюция ТП в разных группах позвоночных животных идет разными путями и обусловлена разными механизмами.

Финансирование: Исследование поддержано грантом Российского научного фонда (проект № 24-24-00480).

The evolution of tandem repeats using the example of two groups of vertebrates

Ostromyshenskii D.I.[1]*, Popova M.A.[1, 2], Travina A.O.[1]

[1] *Institute of Cytology, RAS, St. Petersburg, Russia*

[2] *Center for Molecular and Cellular Biology, Skolkovo Institute of Science and Technology, Moscow, Russia*

** necroforus@gmail.com*

Key words: tandem repeat; heterohromatin; hamster; frog; evalution

Motivation and Aim: Tandem repeats (TR) are an important part of the genome of all vertebrates. Constitutive heterochromatin, which forms such functionally significant parts of chromosomes as centromeric, pericentromeric and subtelomeric regions, consists mainly of TR. TR consist of multiply repeating monomers from several bp up

to several kbp. These monomers form fields up to several tens of mbp long. Specific features of the TR structural organization complicate the assembly, annotation, and mapping of heterochromatic chromosome regions that contain TR as the main component. The evolution of TR is also an extremely poorly studied issue. The emergence of a large number of genome assemblies made using sequencing technologies with long reads (PacBio and OxfordNanopore) makes it possible to more effectively study TR and their evolution.

Methods and Algorithms: We looked for TRs in the genome assemblies of the frog *Rana temporaria* (WGS project CAJIMO01) and hamsters *Cricetulus griseus* (RAZU01), *Mesocricetus auratus* (JAFVMI01), *Phodopus sungorus* (JANBXA01) and *P. roborovskii* (CALSGD01) according to a previously described method [1]. TRs with a field length > 10 kb were considered, except for species of genus *Phodopus* in which TR fields with a length > 5 kb were considered. The founded monomers of TR families were searched in the available genomes of all species of the family Ranidae for the frog (*R. temporaria*, *R. kukuronis*, *R. muscos*a, *Lythobathes sylvaticus*, *L. septentrionalis*, *Aquarana catesbeiana*, *Staurois parvus* and *Glandirana rugosa*) and the subfamily Cricetinae for hamsters (indicated above species). The search was carried out using BLAST with the parameters "dust" "no", "evalue" "1-e4". BLAST hits with a bit-score < 50 and a bit-score/TR monomer length ratio < 1 were removed from further analysis.

Results: 77 TR families were found in the *R. temporaria* genome. We did not find a clearly expressed major TR. 6 TR families are approximately the same in total fields length. Only 15 of the found TR families are species specific. Including three of the most represented in the *R. temporaria* genome Rtem1149A, Rtem138A and Rtem35A. Another 14 TR families were also found in the closely related species *R. kukuronis*, including one of the major families for *R. temporaria* Rtem494A. One of these families, Rtem149A, is present in single copies in the *R. kukuronis* genome, in contrast to the *R. temporaria* genome. In the genomes of all three representatives of Rana, but not of other genera, two TR families Rtem103A and Rtem965A were found, but in *R. kukuronis* and *R. muscosa* these families are represented by single copies. Another major TR of *R. temporaria* Rtem27A is represented by a large number of copies (more than ten thousand) in other *Rana* and single copies in species of other genera. About 10 families of TRs have similar contents in the studied genomes of Ranidae, which may indicate that these TPs originated from interspersed repeats.

In hamsters (Cricetinae), from 45 (*P. roborovskii*) to 65 (*P. sungorus*) TR families were found. In all hamsters, major TR families Cgri33A (*C. griseus*), Maur49A (*M. auratus*) or several major TR families were found – Prob863A, Prob18A and Prob19A (*P. roborovskii*); Psun20A, Psunh1212A, Psun32A (*P. sungorus*). In all studied species, about half of the TR families are species specific and are not found in other species. In most cases, TR families with a relatively small content in the genome are considered non-species-specific. The exception is one of the major TRs of *P. roborovskii* Prob18A, which is found in the genomes of three other species, but in an amount of no more than 100 copies, which is significantly less than the content in the genome of *P. roborovskii*. Other families of TRs with a high content in genomes are either not found in other species or are represented by a small number of copies; such families include, for example, TRs *C. griseus* Cgri72A, Cgri23A; *M. auratus* TR families Maur24A, Maur20A and others. For the genus *Phodopus*, 5 genus-specific TR families were found. Also, in species of this genus, a significant number of species-specific TRs with a monomer about 1 kb and more in length were found. This TRs forming fields more than

100 kbp in length. TRs with such long monomers are often derivatives of interspersed repeats [1, 2], but the TRs we found did not show homology with interspersed repeats.
Conclusion: The most popular model of TR evolution is the library hypothesis [3]. The emergence of genome assemblies of closely related species makes it possible to test this hypothesis on the full set of TRs of different species. It was shown that the distribution pattern of TRs in the genomes of Ranidae generally corresponds to the library hypothesis, while in hamsters it contradicts it. This may indicate that the evolution of TR in different groups of vertebrates follows different paths and is caused by different mechanisms.
Funding: The study is supported Russian Science Foundation (project No. 24-24-00480).

Список литературы/References

1. Komissarov A.S., Gavrilova E.V., Demin S.J., Ishov A.M., Podgornaya O.I. Tandemly repeated DNA families in the mouse genome. *BMC Genomics.* 2011;12:531
2. Ivanova N.G., Stefanova V.N., Ostromyshenskii D.I., Podgornaya O.I. Tandem repeats in the genome of Sus scrofa, their localization on chromosomes and in the spermatogenic cell nuclei. *Russ J Genet.* 2019;55:835-846
3. Fry K., Salser W. Nucleotide sequences of HS-α satellite DNA from kangaroo rat Dipodomys ordii and characterization of similar sequences in other rodents. *Cell.* 1977;12(4):1069-1084

doi 10.18699/bgrs2024-7.1-11

Сравнительная геномика трех видов уток (шилохвость, кряква и обыкновенный гоголь) и эволюция кариотипов у представителей семейства Anatidae (Anseriformes, Aves)

Проскурякова А.[1]*, Беклемишева В.[1], Тишакова К.[1], Романенко С.[1], Андреюшкова Д.[1], Юдкин В.[2, 3], Интересова Е.[2, 4], Янг Ф.[5], Фергюсон-Смит М.[6], Графодатский А.[1]

[1] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[2] *Институт систематики и экологии животных СО РАН, Новосибирск, Россия*

[3] *Новосибирский государственный университет, Новосибирск, Россия*

[4] *Томский государственный университет, Томск, Россия*

[5] *Школа медицины и наук о жизни, Шандонский технологический университет, Цзыбо, Китай*

[6] *Кембриджский ресурсный центр сравнительной геномики, Кембриджский университет, Кембридж, Великобритания*

* *andrena@mcb.nsc.ru*

Ключевые слова: *Anas acuta*; *Anas platyrhynchos*; *Bucephala clangula*; *Burhinus oedicnemus*; авдотка; сравнительный хромосомный пэйнтинг; конститутивный гетерохроматин; рибосомные гены, теломера

Мотивация и цель: Курообразные (Galliformes) и гусеобразные (Anseriformes) – две ветви группы Galloanserae, базальные по отношению к остальным новонебным птицам (Neognathae). В сравнении с курообразными эволюция кариотипов гусеобразных изучена недостаточно, в сравнительные геномные исследования вовлечено небольшое количество видов [1]. Поэтому целью настоящей работы стали сравнительные геномные исследования кариотипов трех представителей отряда гусеобразные из семейства утиные (Anatidae): кряквы (*Anas platyrhynchos*), шилохвости (*Anas acuta*) и обыкновенного гоголя (*Bucephala clangula*) при помощи цитогенетических методов.

Методы и алгоритмы: Для изучения кариотипов трех видов уток были получены первичные клеточные культуры фибробластов и суспензии фиксированных метафазных клеток [2]. Для подсчета количества и выявления морфологии хромосом кариотипы были окрашены красителем Гимза. Также для описания кариотипов были использованы различные методы дифференциального окрашивания (GTG-, CBG-, CDAG-) [3–5]. Сравнение кариотипов трех видов уток осуществлено с помощью метода флуоресцентной *in situ* гибридизации (FISH) с различными типами зондов [6]: содержащих сортированные хромосомы авдотки [7], гены рРНК [8] и теломерные повторы [9].

Результаты: Исследование кариотипов трех видов уток с помощью рутинной окраски и GTG-, CBG- и CDAG-методов показало, что все исследуемые виды имеют одинаковое диплоидное число хромосом ($2n = 80$). Однако кариотип обыкновенного гоголя отличается от кариотипов представителей рода *Anas* увеличенным числом макрохромосом и наличием крупных блоков гетерохроматина на большинстве макроаутосом и на половых хромосомах. Использование зондов, разработанных на основе сортированных хромосом

авдотки, выявило районы гомологии на макро- и микрохромосомах уток. Описанные ранее результаты сравнения кариотипов курицы и авдотки [7] позволили нам установить районы гомологии между хромосомами курицы, референсным геномом птиц и тремя видами уток. Кластеры генов рРНК локализованы на четырех парах микрохромосом шилохвости и на двух парах микрохромосом обыкновенного гоголя. Мы показали, что у видов Anatidae существуют микрохромосомы, обогащенные теломерными последовательностями, что ранее было описано для кариотипов представителей других семейств птиц [10, 11].

Выводы: Впервые получены подробные хромосомные карты для трех видов утиных, содержащие информацию о гомологии с хромосомами авдотки и курицы, распределении конститутивного гетерохроматина, генов рРНК и амплификации теломерных повторов. Результаты межвидовых сравнений показали, что увеличение числа макрохромосом в кариотипе обыкновенного гоголя произошло за счет разрывов двух предковых элементов. Новые данные дополнили информацию о преобразованиях макро- и половых хромосом гусеобразных в ходе эволюции.

Финансирование: Исследование поддержано грантом РНФ (№ 23-74-01016).

Comparative genomic studies in three duck's species (the Northern Pintail, Mallard, and Common Goldeneye) and karyotype evolution in the family Anatidae (Anseriformes, Aves)

Proskuryakova A.[1]*, Beklemisheva V.[1], Tishakova K.[1], Romanenko S.[1], Andreushkova D.[1], Yudkin V.[2, 3], Interesova E.[2, 4], Yang F.[5], Ferguson-Smith M.[6], Graphodatsky A.[1]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russian*

[2] *Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russian*

[3] *Novosibirsk State University, Novosibirsk, Russian*

[4] *Tomsk State University, Tomsk, Russian*

[5] *School of Life Sciences and Medicine, Shandong University of Technology, Zibo, China*

[6] *Cambridge Resource Center for Comparative Genomics, University of Cambridge, Cambridge, UK*

* *andrena@mcb.nsc.ru*

Key words: *Anas acuta*; *Anas platyrhynchos*; *Bucephala clangula*; *Burhinus oedicnemus*; comparative chromosome painting; constitutive heterochromatin; ribosomal gene; telomere

Motivation and Aim: Galliformes and Anseriformes represent two branches of the Galloanserae group, basal to other Neognathae. While Galliformes have been extensively studied, the evolution of Anseriformes karyotypes remains relatively understudied, with only a few species included in comparative genomic analyses [1]. Therefore, this study aims to conduct comparative genomic analyses of the karyotypes of three representatives of the order Anseriformes from the Anatidae family: the mallard (*Anas platyrhynchos*), pintail (*Anas acuta*), and common goldeneye (*Bucephala clangula*) using cytogenetic methods.

Methods and Algorithms: To study the karyotypes of the three duck species, primary cell cultures of fibroblasts and suspensions of fixed metaphase cells were obtained [2]. Giemsa staining was employed to count the number and identify the morphology of chromosomes. Additionally, various differential staining methods (GTG-, CBG-, CDAG-) were utilized to characterize the karyotypes [3–5]. A comparison of the karyotypes of three duck species was carried out using fluorescence *in situ* hybridization (FISH) with different types of probes [6], including stone curlew sorted chromosomes [7], rRNA genes [8], and telomeric repeats [9].
Results: Analysis of the karyotypes of the three duck species using routine staining and GTG, CBG, and CDAG methods revealed that all studied species possess the same diploid number of chromosomes ($2n = 80$). However, the karyotype of the common goldeneye differs from those of representatives of the genus *Anas* due to an increased number of macrochromosomes and the presence of large blocks of heterochromatin on most macroautosomes and sex chromosomes. The utilization of stine curlew chromosome specific probes revealed homologies on duck's macro- and microchromosomes. Previous studies comparing chicken and stone curlew karyotypes [7] facilitated the identification of regions of homology between chicken chromosomes, the reference genome of birds, and the karyotypes of the three duck species. Clusters of rRNA genes were found on four pairs of pintail microchromosomes and on two pairs of common goldeneye microchromosomes. Our study demonstrated that Anatidae species possess microchromosomes enriched in telomeric sequences, a characteristic previously described in the karyotypes of representatives of other avian families [10, 11].
Conclusion: This study represents the first comprehensive chromosomal mapping of three duck species, providing homologies with stone surlew and chicken chromosomes, the distribution of constitutive heterochromatin, rRNA genes, and the amplification of telomeric repeats. Our interspecific comparisons revealed that the increase in macrochromosome numbers in the common goldeneye karyotype is due to fissions in two ancestral elements. These findings contribute to understanding of the evolutionary transformations of macro- and sex chromosomes in Anseriformes.
Funding: This study is supported by the Russian Science Foundation (Grant No. 23-74-01016).

Список литературы/References

1. Degrandi T.M., Barcellos S.A., Costa A.L., Garnero A.D.V., Hass I., Gunski R.J. Introducing the Bird Chromosome Database: An Overview of Cytogenetic Studies in Birds. *Cytogenet Genome Res*. 2020;160(4):199-205. doi 10.1159/000507768
2. Romanenko S.A., Biltueva L.S., Serdyukova N.A. et al. Segmental paleotetraploidy revealed in sterlet (Acipenser ruthenus) genome by chromosome painting. *Mol Cytogenet*. 2015;8(1):90. doi 10.1186/s13039-015-0194-8
3. Seabright M. A rapid banding technique for human chromosomes. *Lancet*. 1971;298(7731):971-972
4. Sumner A.T. A simple technique for demonstrating centromeric heterochromatin. *Exp Cell Res*. 1972;75(1):304-306
5. Lemskaya N.A. et al. A combined banding method that allows the reliable identification of chromosomes as well as differentiation of AT-and GC-rich heterochromatin. *Chromosom Res*. 2018;12(26):307-315. doi 10.1007/s10577-018-9589-9
6. Liehr T. Fluorescence in situ hybridization (FISH). Springer, 2017
7. Nie W. et al. Avian comparative genomics: Reciprocal chromosome painting between domestic chicken (Gallus gallus) and the stone curlew (Burhinus oedicnemus, Charadriiformes). An atypical species with low diploid number. *Chromosom Res*. 2009;17(1):99-113. doi 10.1007/s10577-009-9021-6
8. Maden B.E.H., Dent C.L., Farrell T.E., Garde J., McCallum F.S., Wakeman J.A. Clones of human ribosomal DNA containing the complete 18 S-rRNA and 28 S-rRNA genes. Characterization, a detailed

map of the human ribosomal transcription unit and diversity among clones. *Biochem J.* 1987;246(2):519-527. doi 10.1042/bj2460519
9. Ijdo J.W., Wells R.A., Baldini A., Reeders S.T. Improved telomere detection using a telomere repeat probe (TTAGGG)n generated by PCR. *Nucleic Acids Res.* 1991;19(17):4780. doi 10.1093/nar/19.17.4780
10. Nanda I., Schrama D., Feichtinger W., Haaf T., Schartl M., Schmid M. Distribution of telomeric (TTAGGG)n sequences in avian chromosomes. *Chromosoma.* 2002;111(4):215-227. doi 10.1007/s00412-002-0206-4
11. Delany M.E., Gessaro T.M., Rodrigue K.L., Daniels L.M. Chromosomal mapping of chicken mega-telomere arrays to GGA9, 16, 28 and W using a cytogenomic approach. *Cytogenet Genome Res.* 2007;117:54-63. doi 10.1159/000103165

doi 10.18699/bgrs2024-7.1-12

Филостратиграфический анализ генной сети клеточного цикла у животных

Турнаев И.[1, 2]*, Иванов Р.А.[1], Лашин С.А.[1, 2, 3], Афонников Д.А.[1, 2, 3]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

[3] *Новосибирский государственный университет, Новосибирск, Россия*

* *turn@bionet.nsc.ru*

Ключевые слова: генная сеть; клеточный цикл; филостратиграфический индекс; индекс возрастов; индекс дивергенции

Мотивация и цель: Клеточный цикл (КЦ) – это период существования клетки с момента ее образования путем деления материнской клетки до собственного деления или гибели. Митотический (клеточный) цикл включает четыре последовательно проходимые клеткой фазы: G1 (рост клетки, подготовка к S фазе), S (синтез ДНК), G2 (подготовка к M фазе), M (митоз – разделение материала и структур клетки на две новые клетки и их расхождение). Прохождение клеткой митотического цикла контролируется генной сетью клеточного цикла, включающей центральный транскрипционный фактор, активирующий транскрипцию значительной группы генов КЦ – E2F (E2F1-5); репрессор транскрипции RB (RB1, RBL1, RBL2); комплексы циклин/CDK, управляющие комплексом E2F/RB и через него экспрессией генов КЦ, а также контролирующие другие процессы КЦ; ингибиторы CDK (p21-CDKN1A, p27-CDKN1B и др.); белки, передающие сигналы от ростовых факторов к белкам КЦ (JUN, FOS, MAP-киназы) [1]. Чтобы лучше понять, как изменялась сеть КЦ и функции ее генов в процессе эволюции животных, мы провели анализ возрастов и изменчивости генов с учетом сетевых связей между ними.

Методы и алгоритмы: Данные о генах КЦ и их взаимодействиях были собраны из карточек генов в NCBI gene и информации в статьях PubMed. На основе собранных данных была реконструирована генная сеть клеточного цикла с помощью компьютерной системы Cytoscape [2]. Эволюционные характеристики генов (индекс филостратиграфического возраста (PAI) и индекс дивергенции (DI)) рассчитывали, как описано в [3], с учетом последовательностей ортологичных генов, идентичных рассматриваемому на 50 % и более.

Результаты: Были собраны данные о 131 гене клеточного цикла. Более половины из них (72 гена) вовлечены больше всего в клеточный цикл, 20 % (27 генов) – в транскрипцию РНК, 13 % – в репликацию ДНК, 10 % – в апоптоз, 5 % – в репарацию ДНК, 3 % – в дифференцировку клеток. На основании этой информации с помощью компьютерной системы Cytoscape была реконструирована генная сеть КЦ. Сеть включает 131 ген и 1034 взаимодействия между этими генами: 664 активирующих и 370 подавляющих. В полученной сети каждый ген в среднем взаимодействует с 13,847 генами. Наибольшее число связей с другими генами с сети КЦ имеют гены E2F1 (83 связи), P53 (72), E2F3 (70), E2F2,4,5,6 (по 68 связей), CDK2 (49), c-MYC (48).

Мы провели анализ PAI и DI у генов КЦ. Наиболее древними оказались 12 генов (9 % генов КЦ), произошедших у клеточных организмов. Происхождение 59 % генов сети КЦ соответствует появлению одноклеточных эукариот. Еще 27 % генов появилось с многоклеточными эукариотами, 1 ген – у хордовых (SMAD2), 5 генов (4 % генов КЦ) появились в эволюции у позвоночных, и появление одного гена (NPAT) соответствует появлению костных позвоночных (Euteleostomi).

Наиболее древними генами, появившимися у клеточных организмов, оказались гены, вовлеченные в ДНК репликацию, ДНК репарацию, транскрипцию, синтез фолатов, синтез нуклеотидов. Из 12 генов, появившихся у клеточных организмов, только три вовлечены в первую очередь в клеточный цикл: c-MYC, фосфатаза PP2A и SKP1. Представители семейства транскрипционных факторов E2F E2F2-6 появились в эволюции у одноклеточных эукариот, и только E2F1 появился у многоклеточных организмов. Гены семейства RB явились в эволюции с появлением одноклеточных эукариот – RB1 и RBL1, ген RBL2 – при появлении многоклеточных эукариот. Представители семейства циклинов в основном появились у одноклеточных эукариот и только три циклина (cyclins D2, E1 и E2) – у многоклеточных эукариот, а также самый молодой cyclin I появился в эволюции после появления позвоночных. Почти все гены семейства CDK (циклин-зависимых киназ) произошли у одноклеточных эукариот, за исключением CDK6 и CDK16, появившихся немного позже у многоклеточных эукариот. Интересно, что представители CDI (ингибиторы циклин-зависимых киназ) появляются в эволюции несколько позже. Только два гена CDKN2D (p19-INK4D) и CDKN3 (KAP) произошли у одноклеточных эукариот, гены CDKN1A, CDKN1B, CDKN1C, CDKN2C произошли у многоклеточных эукариот, и два гена CDKN2A (ARF), CDKN2B (p15INK4b) произошли совместно с формированием в эволюции позвоночных.

Все молодые гены КЦ, произошедшие начиная с хордовых (хордовые, позвоночные и косные позвоночные), отличаются низким уровнем дивергенции генов DI (<0.3). Это указывает на то, что данные гены находятся под сильным селективным отбором.

Выводы: Проанализировав публикации PubMed, мы собрали данные по 131 гену клеточного цикла. Анализ возрастов (индекс PAI) генов сети КЦ показал, что представители основных семейств генов КЦ E2F, RB, CDK и циклинов появились в эволюции совместно с образованием одноклеточных или многоклеточных эукариот. Тогда как гены семейства CDI появились несколько позже в основном у многоклеточных эукариот и у позвоночных. Это может быть связано с необходимостью более жесткого контроля КЦ у многоклеточных эукариот и позвоночных.

Финансирование: Исследование поддержано бюджетным проектом № FWNR-2022-0020.

Phylostratigraphic analysis of gene networks of cell cycle in animals

Turnaev I.I.[1, 2]*, Ivanov R.A.[1], Lashin S.A.[1, 2, 3], Afonnikov D.A.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *turn@bionet.nsc.ru*

Key words: gene network; cell cycle; phylostratigraphic age index; divergence index

Motivation and Aim: The cell cycle (CC) is the period of a cell's existence from the moment of its formation by division of the mother cell to its own division or death. The mitotic (cellular) cycle includes four phases that the cell passes through sequentially: G1 (cell growth, preparation for the S phase), S (DNA synthesis), G2 (preparation for the M phase), M (mitosis – separation of cell material and structures into two new cells and their divergence). The mitotic cycle of the cell is controlled by the cell cycle gene network, including the central transcription factor activating the transcription of a significant group of CC genes – E2F (E2F1-5); Rb family (RB1, RBL1, RBL2) are best known as a repressors of the E2F/DP family of transcription factors; cyclin/CDK complexes that control the E2F/RB complex and through it the expression of CC genes, as well as control other CC processes; CDK inhibitors (p21-CDKN1A, p27-CDKN1B and etc.); proteins (JUN, FOS, MAP-kinases) that transmit signals from growth factors to CC proteins [1]. To better understand how the CC network and the function of its genes have changed in evolution, we analyzed the ages and variability of genes with respect to the network connections between them.
Methods and Algorithms: Data on CC genes and their interactions were collected from gene cards in NCBI's Gene and information in PubMed articles. Based on the collected data, the gene network of the cell cycle was reconstructed using the Cytoscape computer system [2]. The evolutionary characteristics of genes ((phylostratigraphic age index (PAI) and divergence index (DI)) were calculated as described in [3], taking into account the sequences of orthologous genes identical to the considered gene by 50 % or more.
Results: Data on 131 cell cycle genes were collected. More than half of them (72 genes) are most involved in the cell cycle, 20 % (27 genes) in RNA transcription, 13 % in DNA replication, 10 % in apoptosis, 5 % in DNA repair and 3 % in cell differentiation. Based on this information, the CC gene network was reconstructed using the Cytoscape computer system. The network includes 131 genes and 1034 interactions between these genes: 664 activating and 370 repressing. In the resulting network, each gene interacts with 13,847 genes on average. E2F1 (83 interactions), P53 (72), E2F3 (70), E2F2,4,5,6 (68 interactions each), CDK2 (49), c-MYC (48) genes have the highest number of interactions with other genes from the CC network.
We analyzed PAI and DI in CC genes. The most ancient were 12 genes (9 % of CC genes) originated in cellular organisms. The origin of 59 % of the genes of the CC network corresponds to the appearance of unicellular eukaryotes. Another 27 % of genes appeared with multicellular eukaryotes, one gene appeared in chordates (SMAD2), 5 genes (4 % of CC genes) appeared in evolution in vertebrates and the appearance of 1 gene (NPAT) corresponds to the appearance of bony vertebrates (Euteleostomi).
The most ancient genes that appeared in cellular organisms were genes involved in DNA replication, DNA repair, transcription, folate synthesis, and nucleotide synthesis. Of the 12 genes that have appeared in cellular organisms, only three are primarily involved in the cell cycle: c-MYC, phosphatase PP2A, and SKP1. Genes of the E2F family of transcription factors E2F2-6 appeared in evolution in unicellular eukaryotes and only E2F1 appeared in multicellular eukaryotes. Genes of the RB family appeared in evolution with the appearance of unicellular eukaryotes – RB1, RBL1, but the RBL2 gene – at the appearance of multicellular eukaryotes. Genes of the cyclin family mainly appeared in unicellular eukaryotes and only 3 cyclins (cyclins D2, E1 and E2) appeared in multicellular eukaryotes, as well as the youngest cyclin I appeared in evolution after

the appearance of vertebrates. Almost all genes of the CDK (cyclin-dependent kinases) family originated in unicellular eukaryotes, with the exception of CDK6 and CDK16, which appeared a little later, in multicellular eukaryotes. Interestingly, genes of CDIs (inhibitors of cyclin-dependent kinases) appear somewhat later in evolution. Only two genes CDKN2D (p19-INK4D) and CDKN3 (KAP) occurred in unicellular eukaryotes, genes CDKN1A, CDKN1B, CDKN1C, CDKN2C occurred in multicellular eukaryotes and two genes CDKN2A (ARF), CDKN2B (p15-INK4b) occurred together with formation in the evolution of vertebrates.

All young CC genes from chordates (chordates, vertebrates, and bony vertebrates) are characterized by a low level of DI gene divergence (< 0.3). This indicates that these genes are under high selective pressure.

Conclusion: Having analyzed Pubmed publications, we collected data on 131 cell cycle genes. Analysis of the ages (PAI index) of the genes of the CC network showed that representatives of the major CC gene families E2F, RB, CDK, and cyclins appeared in evolution together with the formation of unicellular or multicellular eukaryotes. Whereas CDI family genes appeared somewhat later mainly in multicellular eukaryotes and in vertebrates. This may be due to the need for tighter control of CC in multicellular eukaryotes and vertebrates.

Funding: The study is supported by the budget project No. FWNR-2022-0020.

Список литературы/References

1. Johnson D., Walker C. Cyclins and cell cycle checkpoints. *Annu Rev Pharmacol Toxicol*. 1999;39:295-312. doi 10.1146/annurev.pharmtox.39.1.295
2. Shannon P., Markiel A., Ozier O. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-2504
3. Mustafin Z.S. et al. Phylostratigraphic analysis of gene networks of human diseases. *Vavilov J Genet Breed.* 2021;25(1):46-56

doi 10.18699/bgrs2024-7.1-13

Мутация C1473G в гене *Tph2* мыши: от молекулярного механизма до биологических последствий

Хоцкин Н.В., Комлева П.Д., Арефьева А.Б., Базовкина Д.В., Куликов А.В.*
Институт цитологии и генетики СО РАН, Новосибирск, Россия
** v_kulikov@bionet.nsc.ru*

Ключевые слова: серотонин; триптофангидроксилаза 2; мутация; мозг; поведение; мыши

Мотивация и цель: Фермент триптофангидроксилаза 2 (ТПГ2) гидроксилирует L-триптофан до L-5-гидрокситриптофана – первую и лимитирующую стадию синтеза серотонина (5-НТ) в головном мозге млекопитающих [1]. Активность этого фермента является основным фактором, определяющим синтез 5-НТ. Некоторые мутации в гене *Tph2* человека ассоциированы с тяжелыми психопатологиями [2, 3]. Мутация C1473G в гене *Tph2* мыши приводит к замене P447R в молекуле фермента и вдвое снижает активность фермента в мозге [4, 5]. Против мутантной G аллели идет естественный отбор в природных популяциях мышей [6].
Целью исследования является выяснение молекулярного механизма снижения активности мутантного фермента и выяснение биологических последствий данной мутации.
Методы и алгоритмы: Исследования проводили на молодых (3 недели) и взрослых (8 недель) самцах генотипов B6-1473C (дикий тип) и B6-1473G (мутантный тип), полученных скрещиванием их гетерозиготных родителей B6-1473CG. Эти мыши отличались только аллелями C и G. Влияние аллелей на стабильность молекулы фермента оценивали по величине свободной энергии температурной денатурации (dG). Активность ТПГ2, уровни 5-НТ, 5-HIAA определяли с помощью ВЭЖХ, тогда как уровни белка ТПГ2 в мозге определяли с помощью вестерн блота. Влияние C и G аллелей на дистонию задних конечностей, физическое доминирование самцов, привлекательность самцов для самок изучали в тесте подвешивания за хвост, tube test и ольфакторном тесте соответственно. Вероятность передачи аллелей C и G потомкам исследовали в популяционном эксперименте. Активацию врожденного иммунитета проводили однократным введением бактериального липополисахарида (ЛПС, 2 мг/кг, в/б).
Результаты: Мутация C1473G снижала уровень белка ($F(1,15) = 8.84$, $p < 0.01$) и активность ТПГ2 ($F(1,15) = 8.24$, $p < 0.01$) в среднем мозге. Впервые было показано, что аллель G снижает величину dG термической денатурации молекулы ТПГ2: с 1.865 ± 0.064 ккал/моль (дикий тип) до 1.306 ± 0.051 ккал/моль (мутантный белок) ($F(1,15) = 46.7$, $p < 0.001$). Требуется затратить меньше энергии для денатурации мутантного белка, чем белка дикого типа. При одинаковом уровне экспрессии мутантного и дикого типа генов *Tph2* [7] снижение стабильности мутантного белка уменьшает его время жизни, концентрацию и, как следствие, активность.
Взрослые самцы B6-1473C (индекс доминирования, 1.55 ± 0.5) и B6-1473G (индекс доминирования, 1.44 ± 0.5) не различались по способности к физическому доминированию ($F(1,16) < 1$), по привлекательности их мочи для самок ($F(3,76) =$

1.34, $p > 0.05$). Более того, вероятности передачи аллелей C и G потомкам не различались ($\chi^2 = 0.09$, $p > 0.05$).
В то же время у молодых (3 недели) самцов генотипа B6-1473G наблюдалась дистония задних конечностей: в тесте подвешивания за хвост частота ($F(1,41)=7.28$, $p = 0.01$) и длительность ($F(1,41) = 7.44$, $p = 0.009$) скрещивания задних лапок у молодых самцов B6-1473G были выше, чем у молодых самцов B6-1473C. У взрослых самцов B6-1473G дистония исчезала.
Однократное введение 2 мг/кг ЛПС молодым самцам приводит к гибели 42 % самцов генотипа B6-1473G в течение 72 ч после введения, тогда как выживают все самцы генотипа B6-1473C ($p = 0.05$). Введение ЛПС сопровождается значительным увеличением уровня 5-HT ($p < 0.01$) и его основного метаболита, 5-HIAA ($p < 0.05$), в среднем мозге мышей генотипа B6-1473C, но не генотипа B6-1473G через 48 ч после введения токсина.
Выводы: Впервые было показано, что мутация C1473G: 1) снижает стабильность молекулы ТПГ2; 2) вызывает дистонию задних конечностей у молодых мышей; 3) повышает чувствительность молодых мышей к ЛПС. Последние два результата позволяют ответить на ключевой вопрос, почему аллель G не обнаружена в природных популяциях мышей [6]. В природных популяциях ведется естественный отбор против молодых мышей, гомозиготных по аллели G, из-за высокой вероятности дистонии задних конечностей при стрессе, когда они не способны убежать от хищника, и из-за их высокой реакции на стимуляцию врожденного иммунитета, когда они чаще умирают от многочисленных инфекций. Активация 5-HT, по-видимому, необходима для выживания животного при стимуляции иммунитета. Вызванная мутацией низкая активность ТПГ2, по-видимому, не способна обеспечивать необходимую для выживания высокую скорость синтеза 5-HT, что приводит к гибели молодых мышей, гомозиготных по G аллели. Данное исследование впервые позволило проследить цепь событий от молекулярного механизма снижения активности фермента до последствий этого снижения на уровне популяции.
Финансирование: РНФ грант № 24-15-00078.

The C1473G mutation in the mouse *Tph2* gene: from molecular mechanism to biological consequences

Khotskin N.V., Komleva P.D., Arefieva A.B., Bazovkina D.V., Kulikov A.V.*
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** v_kulikov@bionet.nsc.ru*

Key words: serotonin; tryptophan hydroxylase 2; mutation; brain; behaviour; mice

Motivation and Aim: The enzyme tryptophan hydroxylase 2 (TPH2) hydroxylates L-tryptophan to L-5-hydroxy tryptophan – the first and rate-limited reaction of 5-HT synthesis in the mammalian brain [1]. The enzyme activity is the main factor that defines the 5-HT turnover rate in the mammalian brain. Some mutations in the human *Tph2* gene are associated with mental disorders [2, 3]. The C1473G polymorphism in mouse *Tph2* gene results in P447R substitution and two-fold decrease in TPH2 activity in the brain [4, 5]. There is a natural selection process against the G allele in wild mouse populations

[6]. This study aims to clarify the molecular mechanism and the biological consequences of this mutation.

Methods and Algorithms: Experiments were carried out on young (3 weeks old) and adult (8 weeks old) male mice of B6-1473C (wild type) and B6-1473G (mutant) genotypes bred by intercrossing of their B6-1473CG heterozygous parents. These mice differ only in the C and G alleles. The effects of the C and G alleles on the enzyme stability is evaluated by free energy (dG) of its thermal denaturation. TPH2 activity, 5-HT, 5-HIAA levels and TPH2 protein concentration in mouse brain are assayed by HPLC and western blot, respectively. The effects of the C and G alleles on hind legs dystonia, male physical dominance and male attractiveness for females are studied in the tail suspension, tube competition and olfactory tests. The C and G allele transmission probabilities are evaluated in a population experiment. The innate immune system is stimulated by a single administration of bacterial lipopolysaccharide (LPS, 2 mg/kg, ip).

Results: The C1473G mutation decreased TPH2 protein levels ($F(1,15) = 8.84$, $p < 0.01$) and activity ($F(1,15) = 8.24$, $p < 0.01$) in midbrain. For the first time, it was shown that the G allele reduces the dG value of thermal denaturation of the TPH2 molecule: from 1.865 ± 0.064 Kcal/mol (wild-type) to 1.306 ± 0.051 Kcal/mol (mutant) ($F(1,15) = 46.7$, $p < 0.001$). Less energy is needed to denature a mutant than a wild-type protein. At the same level of expression of the mutant and wild type *Tph2* genes [7], the reduced stability of the mutant protein decreases the lifetime, protein concentration and, as a consequence, activity of the enzyme. Adult males B6-1473C (dominance index, 1.55 ± 0.5) and B6-1473G (dominance index, 1.44 ± 0.5) did not differ in their ability to physically dominate ($F(1,16) < 1$), or in the attractiveness of their urine to females ($F(3,76) = 1.34$, $p > 0.05$). Moreover, the probabilities of transmitting the C and G alleles to offspring did not differ ($\chi^2 = 0.09$, $p > 0.05$).

At the same time, in young (3 weeks) males of the B6-1473G genotype, hind limb dystonia was observed: in the tail suspension test, frequency ($F(1,41)=7.28$, $p = 0.01$) and duration ($F(1,41) =7.44$, $p=0.009$) of hind leg crossings were higher in young B6-1473G males than in young B6-1473C males. In adult B6-1473G males, this dystonia disappeared. A single injection of 2 mg/kg LPS to young males leads to the death of 42 % of B6-1473G males within 72 hours after administration, while all B6-1473C males survive ($p < 0.05$). Administration of LPS is accompanied by a significant increase in the level of 5-HT ($p < 0.01$) and its main metabolite, 5-HIAA ($p < 0.05$), in the midbrain of mice of the B6-1473C genotype, but not of the B6-1473G genotype, 48 hours after the toxin administration.

Conclusion: The mutation (1) reduces the stability of the TPH2 molecule, (2) causes hind limb dystonia and (3) increases the sensitivity of young mice to LPS. The last two facts answer the key question: why the G allele is not found in wild mouse populations [6]. In wild populations, natural selection is carried out against young mice homozygous for the G allele: (1) they are not able to escape from a predator due to the stress-induced hind limb dystonia, and (2) they die from numerous infections due to their high response to stimulation of the innate immune system. Activation of the brain 5-HT system appears to be essential for the survival of the animal after immune stimulation. The low activity of TPH2 caused by the mutation seems unable to provide the high rate of 5-HT synthesis necessary for survival, which leads to the death of young mice homozygous for the G allele. The study trace the chain of events from the molecular mechanism of the decrease in enzyme activity to the consequences of this decrease at the population level.

Funding: Russian Science Foundation: grant No. 24-15-00078.

Список литературы/References

1. Walther D.J., Peter J.U., Bashammakh S., Hörtnagl H., Voits M., Fink H., Bader M. Synthesis of serotonin by a second tryptophan hydroxylase isoform. *Science*. 2003;299:76. doi 10.1126/science.1078197
2. Popova N.K., Kulikov A.V. Targeting tryptophan hydroxylase 2 in affective disorder. *Expert Opin Ther Targets*. 2010;14:1259-1271. doi 10.1517/14728222.2010.524208
3. Ottenhof K.W., Sild M., Lévesque M.L., Ruhé H.G., Booij L. TPH2 polymorphisms across the spectrum of psychiatric morbidity: A systematic review and meta-analysis. *Neurosci Biobehav Rev*. 2018;92:29-42. doi 10.1016/j.neubiorev.2018.05.018
4. Zhang X., Beaulieu J.M., Sotnikova T.D., Gainetdinov R.R., Caron M.G. Tryptophan hydroxylase-2 controls brain serotonin synthesis. *Science*. 2004;305(5681):217. doi 10.1126/science.1097540
5. Kulikov A.V., Osipova D.V., Naumenko V.S., Popova N.K. Association between Tph2 gene polymorphism, brain tryptophan hydroxylase activity and aggressiveness in mouse strains. *Genes Brain Behav*. 2005;4(8):482-485. doi 10.1111/j.1601-183X.2005.00145.x
6. Osipova D.V., Kulikov A.V., Mekada K., Yoshiki A., Moshkin M.P., Kotenkova E.V., Popova N.K. Distribution of the C1473G polymorphism in tryptophan hydroxylase 2 gene in laboratory and wild mice. *Genes Brain Behav*. 2010;9(5):537-543. doi 10.1111/j.1601-183X.2010.00586.x
7. Bazhenova E.Y., Fursenko D.V., Kulikova E.A., Khotskin N.V., Sinyakova N.A., Kulikov A.V. Effect of photoperiodic alterations on depression-like behavior and the brain serotonin system in mice genetically different in tryptophan hydroxylase 2 activity. *Neurosci Lett*. 2019;699:91-96. doi 10.1016/j.neulet.2019.01.041

doi 10.18699/bgrs2024-7.1-14

The C886T mutation in the *Th* gene reduces the activity of tyrosine hydroxylase in the brain of mice

Alsalloum I.[1, 2]*, Moskaliuk V.[1], Rakhov I.[1, 2], Kulikov A.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *ismailalsalom7@gmail.com*

Key words: tyrosine hydroxylase; C886T mutation; activity; expression; brain; mice

Motivation and Aim: The dopaminergic (DA) system in the brain plays a pivotal role in regulating the nervous system, endocrine glands, and both adaptive and pathological behaviors [1]. DA synthesis in the brain occurs in two steps from the amino acid L-tyrosine: first, the enzyme tyrosine hydroxylase (TH) converts L-tyrosine to L-3,4-dihydroxyphenylalanine (L-DOPA), and then, the enzyme aromatic amino acid decarboxylase converts L-DOPA to DA. The hydroxylation of L-tyrosine is the key factor determining DA levels in the brain. Some mutations in the human *TH* gene are associated with childhood parkinsonism, an increased risk of Parkinson's disease [2], dystonia [3], and bipolar disorders [4]. Understanding the molecular events associated with mutations in the *TH* gene and their effects on the nervous system, motor function, and psychological functions of humans is restricted due to social and ethical constraints. Therefore, the modeling of molecular events caused by mutations in laboratory rodents becomes highly relevant. The Ensembl genome database (https://www.ensembl.org/ index.html) contains information on 21 SNPs in the mouse *Th* gene, leading to amino acid substitutions in the TH molecule. Among these, only one SNP, the C886T mutation resulting in the R278H substitution in the TH molecule, is identified in widely used laboratory mouse strains. This study aims to investigate the impact of the C886T mutation in the *Th* gene on the TH activity in the mouse midbrain, the region containing the DA neuron bodies.
Methods and Algorithms: This study is carried out on adult males of C57BL/6 (886T, $n = 6$), DBA/2 (886T, $n = 6$), CAST (886C, $n = 5$) lines as well as males and females of F2 segregating intercrosses between C57BL/6 and CAST mice ($n = 42$). TH activity is assayed by HPLC technique and quantified in picomoles (pmol) of L-DOPA synthesized per minute per milligram of protein. Genotyping of the 886T and 886C alleles was accomplished through quantitative real-time qPCR technique. *Th* gene mRNA and TH protein levels are assayed by qPCR and western blot analysis, respectively.
Results: Mice from the three strains exhibited differential TH activity in the midbrain ($F(2,14) = 21.68$, $p < 0.001$). CAST mice displayed significantly higher TH activity compared to C57BL/6 ($p < 0.001$) and DBA/2 ($p < 0.001$) mice. However, no interstrain differences in *Th* gene mRNA levels ($F(2,13) = 2.67$, $p = 0.11$,) or TH protein levels ($F(2,14) = 1.23$, $p = 0.32$) in the midbrain were observed (Fig. 1).
Distribution of TT, TC and CC genotypes among F2 intercross mice was consistent with the expected 1:2:1 ratio ($\chi2(1) = 0.805$, $p > 0.05$) (Table 1).

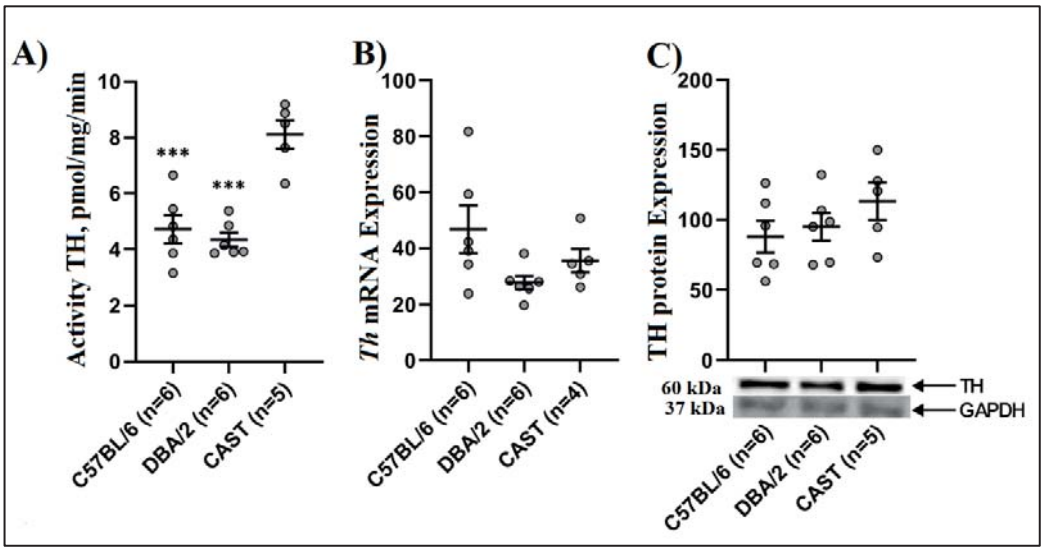


Fig. 1. TH Activity (A), *Th* gene mRNA level (B), and TH protein level (C) in the midbrain of mice of C57BL/6 (TT), DBA/2 (TT), and (CAST) (CC) lines. Individual values are shown with means ± SEM. *Th* gene expression is normalized to *Polr2a* gene expression, and TH protein levels are normalized to GAPDH protein levels. ***$p < 0.001$ vs. CAST. Between-line differences were evaluated using one-way ANOVA

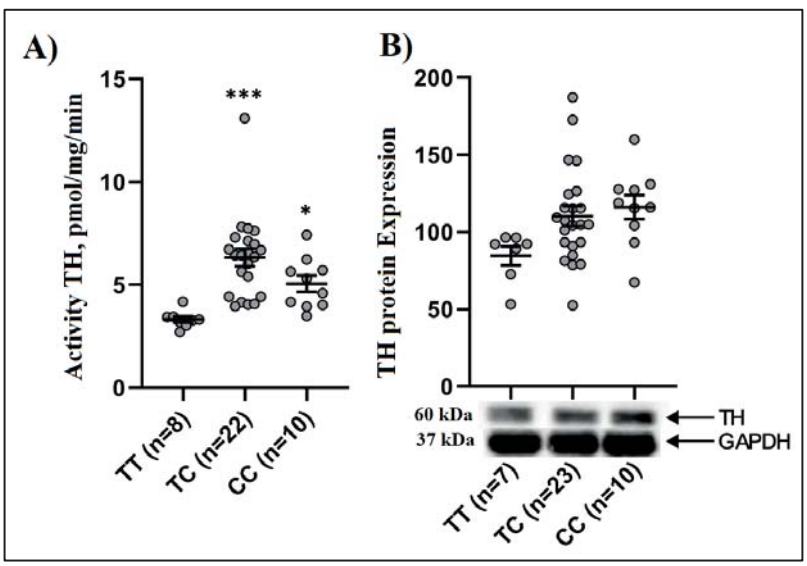


Fig. 2. TH Activity (A) and Protein Level (B) in the midbrain of F2 intercrosses with TT, TC, and CC genotypes. Data for males and females are pooled. Individual values are shown with means ± SEM. The TH protein level is normalized to the GAPDH protein level. *$p < 0.05$, **$p < 0.001$ vs. TT. F2 intercrosses were analyzed using two-way ANOVA. Fisher's LSD method was used for intergroup comparisons

Table 1. Distribution of TT, TC, and CC genotypes among the male and female F2 intercross offspring, obtained from the mating of F1 hybrids (C57BL/6 x CAST)

Genotype	Males	Females	Males + Females
TT	5	3	8
TC	13	10	23
CC	5	6	11

Among F2 intercross mice, TH activity was significantly lower in individuals with the TT genotype compared to TC ($p < 0.001$) and CC ($p = 0.02$) genotypes (Fig. 2A), while TH protein levels showed no significant differences between genotypes ($F(2,38) = 2.22$, $p = 0.12$, Fig. 2B).

Conclusion: This pilot study clearly establishes the significance of the C886T polymorphism as a critical determinant of TH activity in the mouse brain. Notably, being the first naturally occurring common mutation in the mouse *Th* gene shown to impact enzyme activity, this mutation opens up promising avenues for further experimental modeling to explore its effects on physiological functions, both in normal and pathological conditions.

Funding: The study is supported by Russian Science Foundation, grant No. 24-15-00078.

References

1. Channer B., Matt S.M., Nickoloff-Bybel E.A. et al. Dopamine, Immunity, and Disease. *Pharmacol Rev*. 2023;75:62-158. doi 10.1124/pharmrev.122.000618
2. Bademci G., Vance J.M., Wang L. Tyrosine hydroxylase gene: another piece of the genetic puzzle of Parkinson's disease. *CNS Neurol Disord Drug Targets*. 2012;11:469-481. doi 10.2174/187152712800792866
3. Nagatsu T., Nakashima A., Ichinose H., Kobayashi K. Human tyrosine hydroxylase in Parkinson's disease and in related disorders. *J Neural Transm (Vienna)*. 2019;126:397-409. doi 10.1007/s00702-018-1903-3
4. Craddock N., Davé S., Greening J. Association studies of bipolar disorder. *Bipolar Disord*. 2001;3:284-298. doi 10.1034/j.1399-5618.2001.30604.x

doi 10.18699/bgrs2024-7.1-15

The influence of different temperature IVF conditions on development destabilization of CD1 mice

Babochkina T.I.*, Gerlinskaya L.A., Stanova A.K., Sharapova M.B., Shevelev O.B., Moshkin M.P., Moshkin Y.M.

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

** babochkinat@bionet.nsc.ru*

Key words: fluctuation asymmetry (FA); gene expression FA; IVF temperature condition; development destabilization; early development; mouse

Motivation and Aim: The organism development program is determined by the coordinated expression of genes. In multicellular bilateral organisms it is expected that the gene expression in the left/right structures of the body will be symmetrical. However, all stages of the implementation of genetic information are subject to stochastic fluctuations. These stochastic fluctuations may generate fluctuating asymmetry (FA), which can indicate developmental destabilization. Traditionally, this FA can be evaluated on the basis of morphometric measurements of bilateral structures. We proposed to use the evaluation of FA in bilateral structures at the molecular level – at the level of gene expression.

The FA of gene expression in the organism's bilateral structures in a variable environment remains an area of active research. The establishment of the gene expression pattern, including FA, may depend on many environmental factors during early development, including temperature. The temperature can influence the gene expression pattern via epigenetic changes soon after fertilization. This epigenetic modification might be important for the establishment of gene expression patterns in bilateral structures of the organism, which create the FA in adult phenotypes.

The aim of our work is to investigate the effects of different IVF temperature conditions on developmental destabilization (the forming of gene expression patterns, which affect the gene expression FA in bilateral structures) by analyzing the FA gene expression in left\right bilateral organs of adult mice.

Methods and Algorithms: The mice zygotes were subjected to two different developmental temperatures, 35 °C and 37 °C, in the first 24 h after *in vitro* fertilization with subsequent transfer to the recipient female. For the control group, we used *in vivo* fertilization with subsequent transfer to the recipient female. The temperature influence on FA gene expression was studied in 36-week-old adult mice by the RT-PCR method. The 11 – gene "diagnostic set" was used for expression in left\right forelimbs. The data was analyzed via traditional calculation of asymmetry coefficient (K_{FA} =|L-R|\|L+R|) and by an analytical algorithm based on the decomposition of dispersions by the method of principal components (PCA). To study the development destabilization of adult mice we analyzed the FA indicators of gene expression and correlated them with the body weight of mice in aged 3, 8, 18, and 24 weeks.

Results: The weight of mice depends significantly on the IVF temperature conditions and on the age of the mice. The body mass of mice in aged 3, 8, and 18 weeks in the 37 °C group differs from that in the 35 °C group. Also, it was different from the mass of mice in the control group aged 8 weeks.

The levels of FA gene expression are different in the 35 °C and the 37 °C IVF temperature and control groups. The traditional coefficient of symmetry K_{FA} in mice in the 37 °C group was significantly different from that of the mice from the control group. The quantitative estimation of gene expression of 11 genes in the left/right forelimbs of the males by real time PCR and following PCA decomposition demonstrated the left\right forepaw patterns' different gene expression in experimental 35 °C and 37 °C IVF conditions and in control groups.

Conclusion: We propose that the temperature conditions of the period before implantation play a significant role in development, contributing to the fitness of the adult offspring.

Funding: The research was supported by the Russian Science Foundation grant No. 23-14-00179, and carried out using equipment from the Center for Genetic Resources of Laboratory Animals, Federal Research Center ICG SB RAS, supported by the Ministry of Education and Science of Russia (unique project identifier: RFMEFI62119X0023).

doi 10.18699/bgrs2024-7.1-16

Hyperbolic geometry and information acquisition in biological systems

Sharpee T.O.

Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA
Department of Physics, University of California, San Diego, La Jolla, CA, USA
sharpee@salk.edu

Key words: neuroscience; bioinformatics; gene expression; information transmission

Motivation and Aim: Across different scales of biological organization, biological networks often exhibit hierarchical tree-like organization. For networks with such structure, hyperbolic geometry provides a natural metric because of its exponentially expanding resolution. I will describe how the use of hyperbolic geometry can be helpful for visualizing and analyzing information acquisition and learning process from across biology, from viruses, to plants and animals, including the brain.

Methods and Algorithms: We used topological and Bayesian dimensionality reduction methods. These methods were used for embedding distances between data points into different hyperbolic spaces with the goal of finding the best fitting low-dimensional hyperbolic geometry for each dataset.

Results: We find that local noise causes data to exhibit Euclidean geometry on small scales, but that at broader scales hyperbolic geometry becomes visible and pronounced. The hyperbolic maps are typically larger for datasets of more diverse and differentiated cells, e. g. with a range of ages. We find that adding a constraint on large distances according to hyperbolic geometry improves the performance of t-SNE algorithm to a large degree causing it to outperform other leading methods, such as UMAP and standard t-SNE. For neural responses, I will describe data showing that neural responses in the hippocampus have a low-dimensional hyperbolic geometry and that their hyperbolic size is optimized for the number of available neurons. It was also possible to analyze how neural representations change with experience. In particular, neural representations continued to be described by a low-dimensional hyperbolic geometry but the radius increased logarithmically with time. This time dependence matches the maximal rate of information acquisition by a maximum entropy discrete Poisson process, further implying that neural representations continue to perform optimally as they change with experience.

Conclusion: In both gene expression and neural datasets, the low-dimensional geometry expanded logarithmically in time, matching the time of the maximal information that the system could acquire during the time allowed time period. These results suggests new principles for vaccine development and modeling of neural responses.

Funding: This study is supported by AHA-Allen Initiative in Brain Health and Cognitive Impairment award made jointly through the American Heart Association and the Paul G. Allen Frontiers Group (19PABH134610000); National Science Foundation (NSF) grant IIS-1724421; the NSF Next Generation Networks for Neuroscience Program (award 2014217); National Institutes of Health grants U19NS112959 and P30AG068635.

References

1. Zhang H., Rich P.D., Lee A.K., Sharpee T.O. Hippocampal spatial representations exhibit a hyperbolic geometry that expands with experience. *Nat Neurosci*. 2023;26(1):131-139
2. Zhou Y., Smith B.H., Sharpee T.O. Hyperbolic geometry of the olfactory space. *Sci Adv*. 2018;4(8):eaaq1458
3. Zhou Y., Sharpee T.O. Hyperbolic geometry of gene expression. *iScience*. 2021;24(3):102225

doi 10.18699/bgrs2024-7.1-17

Biochemical dynamics of the Siberian wood frog in hypoxia and reoxygenation

Shekhovtsov S.V.[1, 2]*, Bulakhova N.A.[2], Tsentalovich Yu.P.[3], Osik N.A.[3], Meshcheryakova E.N.[2], Poluboyarova T.V.[1], Berman D.I.[2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Institute of the Biological Problems of the North, FEB RAS, Magadan, Russia*

[3] *International Tomography Center, SB RAS, Novosibirsk, Russia*

* *shekhovtsov@bionet.nsc.ru*

Key words: hypoxia; reoxygenation; ^{1}H-NMR; succinate; glycerol; 2,3-butanediol

Motivation and Aim: The Siberian wood frog *Rana amurensis* Boulenger, 1886 is a unique amphibian capable of surviving several months in water with very low oxygen content (up to 0.2 mg/L) [1]. Our aim was to study metabolomic changes of this species at the onset of hypoxia (1 day), as well as in 1 hour of reoxygenation following long-term hypoxia exposure using ^{1}H nuclear magnetic resonance (NMR).

Methods and Algorithms: Adult individuals of *R. amurensis* were collected in the field. One sample was exposed to 1 day of hypoxia in water (n = 6). Another six individuals were kept for 10 days under hypoxia and then returned to normoxic water for 1 h. Organs (liver, heart, and brain) were extracted as quickly as possible, and quantitative metabolomic analysis was performed using ^{1}H NMR according to a procedure described previously [2]. The results were compared to the samples kept under normal oxygen content, as well as those exposed to 30 days of hypoxia [2, 3].

Results: While succinate in most studied vertebrates accumulates under hypoxia and is believed to undergo rapid conversion upon oxygen restoration, our study revealed a decrease in succinate in the brain in reoxygenation, while it remained unchanged in the liver. This observation suggests the existence of a mechanism that inhibits succinate conversion. Furthermore, we observed intriguing disparities concerning two substances with unclear functions: glycerol and 2,3-butanediol. Glycerol exhibited rapid accumulation during hypoxia and equally swift processing during reoxygenation. In contrast, 2,3-butanediol required an extended period to accumulate, yet persisted after reoxygenation.

Conclusion: Of the identified metabolites pool, only a few molecules exhibited continuous dynamics with rapid accumulation or decrease in response to hypoxia and subsequent reversal upon reoxygenation. Certain substances grouped together into unit with similar patterns. To gain a comprehensive understanding of specific substances, additional time points are necessary.

Funding: This study was supported by the Russian Science Foundation (RSF) grant No. 21-74-20050.

References

1. Berman D.I., Bulakhova N.A., Meshcheryakova E.N. The Siberian wood frog survives for months underwater without oxygen. *Sci Rep*. 2019;9(1):13594. doi 10.1038/s41598-018-31974-6
2. Shekhovtsov S.V., Bulakhova N.A., Tsentalovich Y.P. et al. Metabolic response of the Siberian wood frog *Rana amurensis* to extreme hypoxia. *Sci Rep*. 2020;10(1):14604. doi 10.1038/s41598-020-71616-4
3. Shekhovtsov S.V., Bulakhova N.A., Tsentalovich Y.P. et al. metabolomic profiling reveals differences in hypoxia response between Far Eastern and Siberian Frogs. *Animals*. 2023;13(21):3349. doi 10.3390/ani13213349

doi 10.18699/bgrs2024-7.1-18

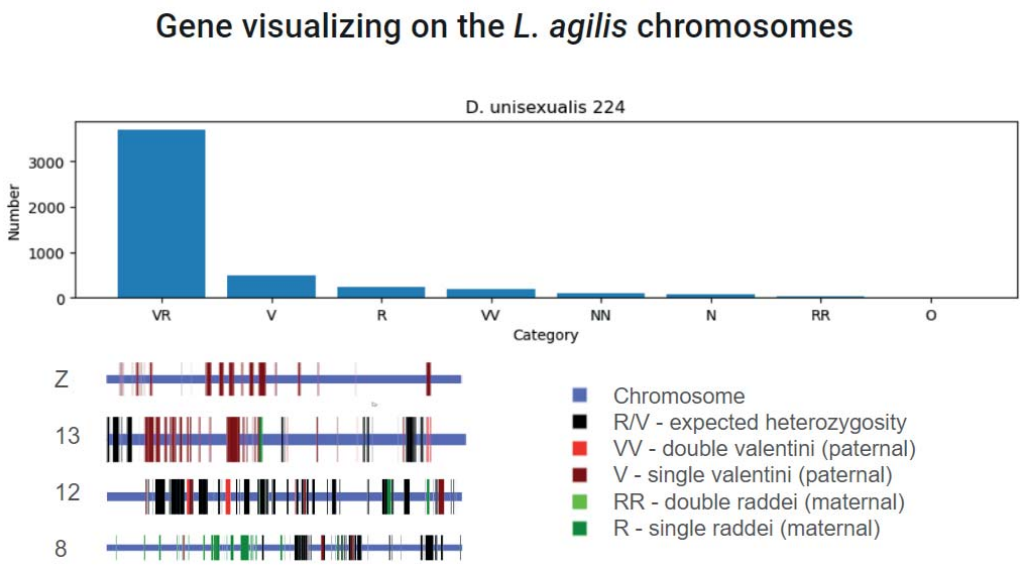
Phylogenetic analysis of the genomes of several sexual and parthenogenetic lizards of the genus *Darevskia*

Urin A.[1, 2]*, Sukhanova X.[1], Korchagin V.[2], Komissarov A.[1, 2], Ryskov A.[2]
[1] *ITMO University, St. Petersburg, Russia*
[2] *Institute of Gene Biology, Moscow, Russia*
* *avel55793@gmail.com*

Key words: partenogenesis, comparative genomic, genome assembling

Motivation and Aim: Parthenogenesis, a form of asexual reproduction where an organism develops from an unfertilized egg, is a rare phenomenon in vertebrates. The genus Darevskia, particularly renowned for its parthenogenetic species, presents a fascinating case study in evolutionary biology due to its complex and reticulated evolutionary history. Recent genomic studies, such as those detailed in the article by Freitas and colleagues [1], highlight the intricate evolutionary pathways that have given rise to these unique reproductive strategies within Darevskia species.
Comparative genomic analyses across various samples and species within this genus can provide significant insights into the mechanisms underlying parthenogenesis and the broader patterns of Squamata evolution. These investigations are crucial not only for understanding the genomic architecture and evolutionary dynamics of Darevskia but also for elucidating the genetic and molecular bases of parthenogenesis. By leveraging high-quality genomic assemblies, researchers can uncover new details about chromosomal evolution, sex determination systems, and reproductive barriers, advancing our knowledge of these rare yet captivating evolutionary phenomena.
Methods and Algorithms: Sequencing results were obtained for the species *Darevskia unisexualis, D. raddei nairensis, D. armeniaca, D. mixta, and D. valentini*. Various samples were sequenced at different times using diverse methods, including Illumina paired-end sequencing, 10x Genomics, and PacBio HiFi sequencing. These methods enabled the assembly of the most complete possible genomes, with PacBio sequencing generating phased haplotypes. Assemblers such as Canu, Flye, and Hifiasm were utilized for the genome assemblies [2].
To assess sample heterozygosity, k-mer analysis was performed. For comparing genome and sex chromosome evolution, conserved genomic loci from the odb_10 [3] dataset were selected and compared within and between genomes. This was accomplished through alignment of the obtained sequences using ClustalW and evaluating the distance matrices using the k-nearest neighbors algorithm (KNN) method. The closely related genome of *Lacerta agilis* was used for visualizing gene locations
Results: Genome assemblies were obtained for three samples of *Darevskia unisexualis*, three of *D. raddei nairensis*, two of *D. valentini,* one of *D. mixta*, and one of *D. armeniaca*. The assemblies demonstrated good results in key quality metrics such as N50, L50, and BUSCO scores. K-mer analysis revealed that the genomes of *D. armeniaca and D. unisexualis* are predominantly heterozygous, while the other species exhibited significantly lower heterozygosity. An exception was one sample of *D. raddei nairensis,* which also showed a high level of heterozygosity. Bioproject access numbers PRJNA1029803, PRJNA1024511.

Comparative analysis of conserved genes indicated that one sample of *D. raddei* exhibited hybrid origins across much of its genome, deriving from different phylogenetic branches within the genus. Additionally, several loci, expected to be autosomal and present in two copies in the diploid genome based on the outgroup, were found to be in hemizygous state in all *D. raddei* and *D. unisexualis* samples, but not in *D. valentini.*
The figure illustrates gene categories for the entire genome and highlights several specific loci. On chromosome Z, hemizygosity was expected, while chromosome 12 showed predominantly heterozygous loci. In contrast, chromosomes 13 and 8 presented unexpected hemizygosity, demonstrating intriguing variations in gene copy number and suggesting complex genomic evolution in these species.



Conclusion: High-quality draft genomes have been obtained, which can serve as valuable resources for various studies.These reliable results can be utilized by molecular biologists and geneticists for research into chromosomal evolution, sex determination, and reproductive mechanisms.
The heterozygosity observed in one sample of *D. raddei nairensis* is particularly intriguing, suggesting that this population might be closer to *D. unisexualis* than to other *D. raddei* samples. However, to confirm the hypothesis of increased heterozygosity and to better understand the genomic composition, further analysis of additional samples from this population is necessary.
Additionally, the varying ploidy observed in certain loci suggests that these regions may have been translocated from autosomes to sex chromosomes. This discovery highlights the dynamic nature of genome evolution in Darevskia species and underscores the need for ongoing genomic studies to fully elucidate these complex processes.
Funding: This research was funded by the Russian Science Foundation (RSF) Research Project No. 19-14-00083.

References

1. Freitas S., Westram A.M., Schwander T. et al. Parthenogenesis in Darevskia lizards: A rare outcome of common hybridization, not a common outcome of rare hybridization. *Evolution*. 2022;76(5):899-914. doi 10.1111/evo.14462
2. Ochkalova S., Korchagin V., Vergun A., Urin A., Zilov D., Ryakhovsky S., Girnyk A., Martirosyan I., Zhernakova D.V., Arakelyan M., Danielyan F. First genome of rock lizard Darevskia valentini involved in formation of several parthenogenetic species. *Genes*. 2022;13:1569. https://doi.org/10.3390/genes13091569
3. Kuznetsov D., Tegenfeldt F., Manni M., Seppey M., Berkeley M., Kriventseva E.V., Zdobnov E.M. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res*. 2023;51(D1):D445-D451. doi 10.1093/nar/gkac998

7 Симпозиум «Генетика/ геномика, биоинформатика и системная биология животных»

Symposium “Genetics/genomics, bioinformatics and systems biology of animals”

doi 10.18699/bgrs2024-7.2-01

Потенциальная роль белка BMAL1 в развитии гипомиелинизации и аффективных расстройств

Амстиславская Т.Г.[1, 2]*, Смирнова К.В.[1, 2]**, Бородина С.О.[3], Ярных В.Л.[4], Смирнова Л.П.[5]

[1] *НИИ нейронаук и медицины, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Сибирский государственный медицинский университет, Томск, Россия*

[4] *НИИ биологии и биофизики Томского государственного университета, Томск, Россия*

[5] *НИИ психического здоровья Томского НИМЦ, Томск, Россия*

** amstislavskayatg@neuronm.ru; ** vedelina@mail.ru*

Ключевые слова: disc1; bmal1; депрессия; гипомиелинезация

Мотивация и цель: Ген *DISC1* играет ключевую роль в работе нервной системы (НС). Белок DISC1 регулирует пролиферацию и дифференцировку нейрональных стволовых клеток, миграцию и созревание нейронов, формирование их синапсов, транспорт веществ, клеточных компонент и работу сигнальных каскадов внутри клетки [1, 2]. Менее изучена роль DISC1 в процессах глиагенеза, однако известно, что он участвует в развитии астроцитов, стимулируя астрогенезис, и олигодендроцитов, определяя правильное время и место дифференцировки их предшественников [3, 4]. При таком разнообразии функций мутации в гене *Disc1* являются ключевым предиктором развития патологических состояний, связанных с нарушениями нейроразвития, синаптической передачи, внутриклеточного сигналинга и циркадных ритмов, например, при психиатрических, нейродегенеративных и демиелинизирующих заболеваниях [5]. Недавно также показано участие DISC1 в циркадных процессах, которые критически важны для функционирования НС. У мышей с нокаутом гена *Disc1* происходит ускоренная деградация циркадного белка BMAL1 за счет усиления активности фермента гликоген синтазкиназы-3 (GSK-3), непосредственно входящего в интерактом DISC1 [6]. Помимо циркадных функций, BMAL1 играет ключевую роль в развитии и работе НС [7]. В мозге взрослых мышей показана высокая экспрессия *Bmal1* как в нейронах, так и в глиальных клетках, варьирующая в разных областях мозга. В том числе BMAL1 играет важную роль в процессах миелинезации, и установлено, что нарушение ритма экспрессии *Bmal1* в предшественниках олигодендроцитов CA1 области гиппокампа и префронтальной коры приводит к развитию депрессивноподобного и тревожного поведения [8, 9]. Характерно, что у мышей с точечной мутацией Q31L в гене *Disc1*, проявляющих выраженный депрессивноподобный фенотип, выявлено снижение взаимодействия DISC1 и GSK-3, которое приводит к усилению энзиматической активности этого фермента [10–12]. Таким образом, изучение распределения белка BMAL1 в ключевых областях мозга, вовлеченных в регуляцию поведения у мышей с мутацией Q31L в гене *Disc1*, поможет выявить новые механизмы патогенеза аффективных расстройств.

Методы и алгоритмы: В работе использовались самцы мышей в возрасте 2–4 месяцев двух генетических линий: C57BL/6 (далее WT) и гомозиготные мыши с точечной мутацией в гене *Disc1*: Disc1-Q31L$^{-/-}$ (Q31L), по 15 мышей на группу.

Животные были получены из уникальной научной установки «Биологическая коллекция – генетические биомодели нейропсихических заболеваний» (№ 493387) Научно-исследовательского института нейронаук и медицины г. Новосибирска [13]. Мышей содержали в соответствии со стандартными требованиями ухода за лабораторными животными согласно руководству ГОСТ 33216-2014, при цикле освещенности 12:12 часов. Тестирование проводили по стандартным протоколам в следующем порядке: открытое поле, тест социальных предпочтений, тест принудительного плавания. Автоматический трекинг и сбор данных осуществляли с помощью оборудования и программного обеспечения Ethovision XT10 (Noldus International Technology). Иммуногистохимический анализ (ИГХ) проводили на замороженных срезах мозга по стандартной методике с использованием первичных антител против BMAL1 (1:800, NovusBio, NB100-2288SS) и вторичных – Alexa fluor 488 (1:500, ThermoFisher scientific, A-11008).

У больных аффективными расстройствами, проходивших лечение в отделении аффективных состояний НИИ психического здоровья ТНИМЦ, оценивали содержание миелина в головном мозге с использованием метода быстрого МПФ-картирования. У пациентов были диагностированы биполярно-аффективное (БАР) и рекурентное депрессивное расстройство (РДР); средний возраст больных 37.55±15.58 года, длительность заболевания 11 [3.5; 16.5] лет, возраст пациентов к началу заболевания 22 [19; 31.5] лет. Было проведено магнитно-резонансное исследование головного мозга с использованием 1.5 Т МР-сканера (Ingenia Evolution, Philips, Netherland) с напряженностью магнитного поля 1.5 Т и амплитудой градиентов 45 мТ/м с использованием 20-канальной цифровой катушки для головы и шеи. Для картирования МПФ (молекулярной протонной фракции) применяли быстрый одноточечный протокол с синтетическим референтным изображением, аналогичный использованному ранее [14, 15]. Карты МПФ были реконструированы с помощью специального программного обеспечения, доступного по адресу https://www.macromolecularmri.org/. Для определения уровня миелинизации головного мозга использовали программу RadiAnt DICOM Viewer (64-bit).

Все полученные данные подвергали статистическому анализу с использованием программного обеспечения Jamovi для Windows. Для выборок с нормальным распределением данных использовали однофакторный дисперсионный анализ ANOVA для независимых выборок, для зависимых выборок – ANOVA для повторных измерений. Post-hoc анализ проводили с применением критерия Фишера (LSD), где значения $p < 0.05$ указывают на статистически значимую разницу [16]. Для независимых выборок, распределение которых не соответствует нормальному, проводили тест Краскела–Уоллиса с последующим применением критерия Манна–Уитни (M-U) для попарных сравнений.

Результаты: Мыши с мутацией Q31L в сравнении с мышами дикого типа (WT) проявляют признаки депрессивноподобного фенотипа, что выражается в статистически значимом увеличении времени дрейфа в тесте принудительного плавания ($F[1; 9.44] = 5.83$, $p < 0.05$) и снижении социальной мотивации ($F[1; 16] = 6.51$, $p < 0.05$), без изменения социально обусловленного предпочтения места и проявления тревожности. Показано, что мутация Q31L значимо снижает иммунофлуоресценцию белка BMAL1 в CA1 области гиппокампа ($F[1; 7] = 35.0$, $p < 0.001$) и латеральной хабенуле ($F[1; 8.51] = 35.7$, $p < 0.001$). При этом не было выявлено изменений в иммунофлуоресценции BMAL1 в других областях

гиппокампа, медиальной хабенуле, миндалевидном комплексе и супрахиазматическом ядре мозга мышей.
Выявленные в исследованиях на мышах области головного мозга, показавшие снижение экспрессии белка BMAL1, могут указывать на снижение миелинизации в этих областях, сочетающееся с развитием депрессивноподобного поведения. И действительно, в проведенных исследованиях на больных рассеянным склерозом выявлено значимое снижение миелинизации, оцененное с помощью МПФ. В предварительных исследованиях изменения миелинизации у больных с БАР и РДР выявлено значимое (с высокой степенью достоверности $p < 0.0001$) снижение миелинизации, оцененное с помощью МПФ, белого вещества головного мозга больных. Процесс гипомиелинизации является диффузным, так как был значимо выражен во всех отделах больших полушарий головного мозга у пациентов с аффективными расстройствами, включая перивентрикулярную область, область базальных ядер (скорлупа и таламус), в сравнении со здоровыми лицами ($p < 0.05$). Эти результаты подтверждают роль подкорковых структур (и в частности таламуса) в формировании симптомов депрессивного расстройства и депрессивной фазы БАР, механизм развития которых может быть объяснен наличием проекционной связи между парафасцикулярными субпопуляциями таламуса и прилежащего ядра, хемогенетическое торможение которого вызывало депрессивное состояние у мышей [17].
Выводы: Таким образом, данные результаты указывают на наличие связи между гипомиелинизацией головного мозга и развитием депрессивноподобного поведения у мышей. У обследованных больных с диагностированными аффективными расстройствами БАР и РДР выявлено значимое, с высокой степенью достоверности, снижение миелинезации белого вещества головного мозга в сравнении со здоровыми лицами. В результате обобщения экспериментальных и литературных данных мы предполагаем, что с высокой степенью вероятности основным действующим фактором в развитии этих изменений является белок BMAL1.
Финансирование: Исследование поддержано грантом РНФ № 23-75-00023.

Potential role of BMAL1 protein in the development of hypomyelination and affective disorders

Amstislavskaya T.G.[1, 2]*, Smirnova K.V.[1, 2]**, Borodina S.O.[3], Yarnykh V.L.[4], Smirnova L.P.[5]

[1] *Research Institute of Neuroscience and Medicine, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Siberian State Medical University, Tomsk, Russia*
[4] *Research Institute of Biology and Biophysics, Tomsk State University, Tomsk, Russia*
[5] *Research Institute of Mental Health, Tomsk National Research Medical Center, Tomsk, Russia*
** amstislavskayatg@neuronm.ru; ** vedelina@mail.ru*

Key words: disc1; bmal1; depression; hypomyelination

Motivation and Aim: The *DISC1* gene plays a key role in the nervous system (NS) functioning. The DISC1 protein regulates proliferation and differentiation of neuronal

stem cells, neurons migration and maturation, synapses formation and maintenance, transport of substances and cellular components, also involved in signaling cascades within cells [1, 2]. Less studied DISC1 functions in gliagenesis, anyway it is known DISC1 stimulate astrogenesis and determine time and place of the oligodendrocyte progenitors differentiation [3, 4]. With such variety of functions, mutations in the *Disc1* gene are a key predictor of pathological conditions associated with neurodevelopmental disorders, disruption of synaptic transmission, intracellular signaling and circadian rhythms, for example, in psychiatric, neurodegenerative and demyelinating diseases [5]. Recently was shown involvement of DISC1 in circadian processes which are critically important for NS functioning. *Disc1*-knockout in mice leads to accelerated degradation of BMAL1 circadian protein through the increased activity of glycogen synthase kinase-3(GSK-3) which is an DISC1 interacting protein [6]. Besides circadian rhythms, BMAL1 play a key role in the NS development and functioning [7]. In the adult mice brain was shown high *Bmal1*expression in neurons and glial cells. Especially BMAL1 play an important role in the myelinisation and it is determined that disrupted rhythm of *Bmal1* expression in hippocampal CA1 and prefrontal cortex oligodendrocyte progenitor cells leads to depressive like and anxious behavior [8, 9]. Interesting that mice with Q31L point mutation in *Disc1* gene characterized with decreased DISC1-GSK-3 interaction and increased enzymatic activity of GSK-3 [10–12]. Thus study of BMAL1 abundance in the key brain structures involved in Q31L mice behavior may provide new mechanisms of mood disorders pathogenesis.

Methods and Algorithms: In that work male mice aged 2–4 months of two genetic lines: C57BL/6 (hereinafter WT) and homozygous mice with a point mutation in the *Disc1* gene: Disc1-Q31L$^{-/-}$ (Q31L) were used in 15 mice per group. Animals were obtained from the unique scientific guideline "Biological collection – genetic biomodels of neuropsychiatric diseases" (No. 493387) of the Research Institute of Neuroscience and Medicine in Novosibirsk [13]. Mice were housed according standard requirements for the care of laboratory animals according to the guidelines of GOST 33216-2014, with a light cycle of 12:12 hours. Testing was carried out according to standard protocols in the following order: open field, social preference test, forced swimming test. Automatic tracking and data collection were carried out using Ethovision XT10 hardware and software (Noldus International Technology). Immunohistochemical analysis (IHC) was performed on frozen brain sections using standard methods with primary antibodies against BMAL1 (1:800, NovusBio, NB100-2288SS), and secondary antibodies – Alexa fluor 488 (1:500, ThermoFisher scientific, A-11008).

In patients with affective disorders who were treated in the department of affective states of the Research Institute of Mental Health TNRMC, the myelin content in the brain was assessed using the fast MPF mapping method. The patients were diagnosed with bipolar affective disorder (BD) and recurrent depressive disorder (RDD), with an average age of patients 37.55±15.58 years, disease duration 11 [3.5; 16.5] years and the age of the patients at the onset of the disease was 22 [19; 31.5] years. MRI data were acquired on a 1.5 T clinical scanner (Ingenia Evolution, Philips, Netherland), 45 mT/m gradient amplitude, using a 20-channel digital head and neck coil. The fast MPF mapping protocol was implemented using a standard manufacturer's 3D spoiled gradient-echo sequence according to the single-point synthetic reference method [14, 15]. MPF maps were reconstructed using custom C++ language software (available at https://www.macromolecularmri.org/) based on the single-point synthetic reference

algorithm. The level of brain myelination was determined using software RadiAnt DICOM Viewer (64-bit).

All obtained data were subjected to statistical analysis using Jamovi software for Windows. For samples with normal data distribution, one-way ANOVA analysis of variance was used for independent samples; for dependent samples – repeated measures ANOVA. Post-hoc analysis was performed using Fisher's test (LSD), where $p < 0.05$ indicate a statistically significant difference [16]. For independent samples not normally distributed, the Kruskal–Wallis test was performed, followed by the Mann–Whitney (M-U) test for pairwise comparisons.

Results: Mice with the Q31L mutation, compared to wild-type WT, show signs of a depressive-like phenotype, which is expressed significant floating time increase in the forced swim test ($F[1;9.44] = 5.83$, $p < 0.05$) and a decrease of social motivation ($F[1; 16] = 6.51$, $p < 0.05$), without changes in social conditioned place preference and anxiety. It was shown that the Q31L mutation significantly reduces the immunofluorescence of the BMAL1 protein in the CA1 region of the hippocampus ($F[1;7] = 35.0$, $p < 0.001$) and the lateral habenula ($F[1;8.51] = 35.7$, $p < 0.001$). However, no changes were detected in BMAL1 immunofluorescence in other areas of the hippocampus, medial habenula, amygdala complex and suprachiasmatic nucleus of the mouse brain. The areas of the brain identified in mouse studies that showed decreased expression of the BMAL1 protein may indicate decreased myelination in these areas, combined with the development of depressive-like behavior. Indeed, studies conducted on patients with multiple sclerosis have revealed a significant decrease in myelination, assessed using MPF. Preliminary studies of changes in myelination in patients with bipolar disorder and RDD revealed a significant ($p < 0.0001$) decrease in myelination the white matter of the brain, assessed using MPF. The process of hypomyelination is diffuse, as it was significantly expressed in all parts of the cerebral hemispheres in patients with affective disorders, including the periventricular region, the region of the basal ganglia (putamen and thalamus), in comparison with healthy individuals ($p < 0.05$). These results confirm the role of subcortical structures (and in particular the thalamus) in the formation of symptoms of depressive disorder and the depressive phase of bipolar disorder. The mechanism of development of these disorders can be explained by the presence of a connection between subpopulations of the thalamus and nucleus accumbens, the chemogenetic inhibition of which caused a depressive state in mice [17].

Conclusion: Thus, these results indicate a link between brain hypomyelination and the development of depressive-like behavior in mice. In the examined patients with diagnosed affective disorders: bipolar disorder and RDD, a significant, with a high degree of reliability, decrease in myelination of the white matter of the brain was revealed in comparison with healthy individuals. As a result of a generalization of experimental and literary data, we assume that with a high degree of probability, the main active factor in the development of these changes is the BMAL1 protein.

Funding: The study is supported the Russian Science Foundation No. 23-75-00023.

Список литературы/References

1. Tomoda T., Hikida T., Sakurai T. Role of DISC1 in Neuronal Trafficking and its Implication in Neuropsychiatric Manifestation and Neurotherapeutics. *Neurotherapeutics*. 2017;14(3):623-629
2. Tropea D. et al. Mechanisms underlying the role of DISC1 in synaptic plasticity. *J Physiol.* 2018;596(14):2747-2771
3. Wang S. et al. DISC1 regulates astrogenesis in the embryonic brain via modulation of RAS/MEK/ERK signaling through RASSF7. *Development*. 2016;143(15):2732-2740. doi 10.1242/dev.133066

4. Katsel P. et al. Overexpression of truncated human DISC1 induces appearance of hindbrain oligodendroglia in the forebrain during development. *Schizophr Bull*. 2018;44(3):515-524
5. Rusznák Z. et al. Adult neurogenesis and gliogenesis: possible mechanisms for neurorestoration. *Exp Neurobiol*. 2016;25(3):103-112
6. Lee S.B. et al. Disrupted-in-schizophrenia 1 enhances the quality of circadian rhythm by stabilizing BMAL1. *Transl Psychiatry*. 2021;11(1):110
7. Zheng Y. et al. Neural function of Bmal1: an overview. *Cell Biosci*. 2023;13(1):1
8. Huang S. et al. Demyelination regulates the circadian transcription factor Bmal1 to signal adult neural stem cells to initiate oligodendrogenesis. *Cell Rep*. 2020;33(7):108394
9. Zuo Y. et al. Circadian misalignment impairs oligodendrocyte myelination via Bmal1 overexpression leading to anxiety and depression-like behaviors. *J Pineal Res*. 2024;76(1):e12935
10. Lipina T.V. et al. Synergistic interactions between PDE4B and GSK-3: DISC1 mutant mice. *Neuropharmacology*. 2012;62(3):1252-1262
11. Clapcote S.J. et al. Behavioral phenotypes of disc1 missense mutations in mice. *Neuron*. 2007;54(3):387-402
12. Lipina T.V. et al. Genetic and pharmacological evidence for schizophrenia-related Disc1 interaction with GSK-3. *Synapse*. 2011;65(3):234-248
13. Petrova E. et al. Maintenance of genetically modified mouse lines: input to the development of bio-collections in Russia. *Lab Anim Sci*. 2018;2. doi 10.29296/2618723X-2018-02-01
14. Yarnykh V.L. Time-efficient, high-resolution, whole brain three-dimensional macromolecular proton fraction mapping. *Magn Reson Med*. 2016;75(5):2100-2106
15. Smirnova L.P. et al. Global hypomyelination of the brain white and gray matter in schizophrenia: quantitative imaging using macromolecular proton fraction. *Transl Psychiatry*. 2021;11(1):365
16. Azizi F., Ghasemi R., Ardalan M. Two common mistakes in applying ANOVA test: Guide for biological researchers. Available ft *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4221619
17. Zhang Y. et al. Targeting thalamic circuits rescues motor and mood deficits in PD mice. *Nature*. 2022;607(7918):321-329

doi 10.18699/bgrs2024-7.2-02

Фенотипическая характеристика новой ENU-мутантной линии мышей с аудиогенной эпилепсией – Сократ

Кулдаева В.П.[1]*, Гавриш М.С.[1], Бабаев А.А.[1], Кондакова Е.В.[1], Борисова Е.В.[2], Тарабыкин В.С.[2]

[1] *Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия*

[2] *Institute of Cell Biology and Neurobiology, Charité-Universitätsmedizin, Берлин, Германия*

** verunya.rubackova@mail.ru*

Ключевые слова: эпилепсия; ENU-индуцированный мутагенез; развитие коры головного мозга

Мотивация и цель: Эпилепсия – неврологическое заболевание с многопрофильным и сложным патогенезом, характеризующееся повторяющимися спонтанными припадками, вызванными аберрантным синхронизированным возбуждением нейронов из-за дисбаланса между возбуждающими и тормозными компонентами центральной нервной системы (ЦНС) [1, 2]. Важная роль в развитии эпилептогенеза отведена генетическим факторам, что подтверждается проведением множества исследований по изучению генетической детерминированности эпилепсий [3]. Для исследования нейрофизиологических, патофизиологических и генетических основ развития эпилептогенеза, а также для создания принципиально новых методов коррекции эпилептической активности широко используются животные модели заболевания. Выявление и характеристика новых мутантных мышей с последующей идентификацией и характеристикой генов, отвечающих за мутации, позволят значительно улучшить наше понимание генетической регуляции развития и последующих нарушений в работе коры головного мозга.

Методы и алгоритмы: Для создания принципиально новых линий мышей был использован химический ENU-направленный мутагенез. Отбор мышей, проявляющих склонность к эпилептиформной активности, проводился при помощи моделирования аудиогенных судорог. За основу классического фенотипического описания новой мышиной модели эпилепсии было взято поведенческое тестирование – тест «открытое поле», тест на условную реакцию пассивного избегания (УРПИ) и протокол измерения акустической реакции вздрагивания. Было проведено SNP-картирование для выявления генов-мишеней, изменяющих свою экспрессию в мышах, проявляющих эпилептическую активность; *in situ* гибридизация проводилась для выявления локализации мРНК, отобранных по результатам секвенирования генов-кандидатов в коронарных срезах головного мозга; иммуногистохимическое окрашивание применялось для характеристики цитоархитектуры коры и гиппокампа мутантных мышей.

Результаты: С помощью ENU-направленного мутагенеза была создана новая мутантная линия мышей, обладающая эпилептиформной активностью. Используя метод моделирования аудиогенных судорог (методика Крушинского [4]), был произведен отбор мышей со склонностью к эпилептическим припадкам, в результате чего была выведена принципиально новая линия мышей – Socrates. Проведенные исследования с применением поведенческих тестов для описания

фенотипа этих мутантных мышей показали, что особи линии Soc обладают пониженной двигательной активностью и способностью к обучению и при этом повышенной ориентировочно-исследовательской активностью, что коррелирует с их повышенной возбудимостью. Применение молекулярно-генетических методов по разработке панели SNP позволило провести картирование участков ДНК для выявления генов-мишеней, участвующих в формировании эпилепсии. Реализованный в дальнейшем транскриптомный анализ генов-мишеней показал изменения экспрессии этих генов. Изучение паттернов экспрессии генов, кодирующих протеинкиназы, ГАМК- и глутаматных рецепторов и мембранных ионных путей показало снижение экспрессии, что способствует нарушению нейротрансмиссии в коре головного мозга мышей Soc, в конечном итоге являясь одной из причин формирования эпилептической активности. При изучении цитоархитектуры коры и гиппокампа было обнаружено увеличение количества интернейронов в коре и увеличение количества астроцитов в гиппокампе, что свидетельствует о нарушении состояния баланса возбуждения/торможения у мутантных особей линии Soc, что также является одной из причин, формирующих эпилептиформное состояние.

Выводы: В этом исследовании нам удалось не только создать и охарактеризовать новую линию мышей с повышенной склонностью к эпилептиформной активности, но и выявить некоторые молекулярно-генетические и морфофизиологические, а также цитоархитектурные изменения, приводящие к изменению баланса между возбуждением и торможением в нейрональной сети мутантных мышей.

Финансирование: Исследование проводилось в рамках государственного задания Минобрнауки России (FSWR-2023-0029).

Phenotypic characterization of novel ENU-induced mouse mutant with audiogenic epilepsy – Socrates

Kuldaeva V.P.[1]*, Gavrish M.S.[1], Babaev A.A.[1], Kondakova E.V.[1], Borisova E.V.[2], Tarabykin V.S.[2]

[1] *Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia*

[2] *Institute of Cell Biology and Neurobiology, Charité-Universitätsmedizin, Berlin, Germany*

* *verunya.rubackova@mail.ru*

Key words: epilepsy; ENU-induced mutagenesis; development of the cerebral cortex

Motivation and Aim: Epilepsy is a neurological disease with a multidisciplinary and complex pathogenesis, characterized by recurrent spontaneous seizures caused by aberrant synchronized neuronal firing due to an imbalance between excitatory and inhibitory components of the central nervous system (CNS) [1, 2]. An important role in epileptogenesis is assigned to genetic factors, which are confirmed by many studies on the genetic determination of epilepsy [3]. Animal models of the disease are widely used to study the neurophysiological, pathophysiological, and genetic basis of epileptogenesis, as well as to create fundamentally new methods for correcting epileptic activity. Identification and characterization of new mutations in mice, followed by identification and characterization of the genes responsible for the mutations, will

significantly improve our understanding of the genetic regulation of corticogenesis and subsequent developmental disorders of the cerebral cortex.

Methods and algorithms: To create fundamentally new mouse strains, chemical ENU-directed mutagenesis was used. The selection of mice prone to epileptiform activity was carried out by modeling audiogenic seizures. The classical phenotypic description of the new mouse model of epilepsy was based on behavioral testing, including the "Open Field" test, the conditioned passive avoidance response test (CPAT), and the protocol for measuring the acoustic startle response. SNP mapping was performed to identify target genes that change their expression in mice exhibiting epileptic activity. In situ hybridization was conducted to identify the localization of mRNAs selected from candidate gene sequencing results in coronal brain sections. Immunohistochemical staining was utilized to characterize the cytoarchitecture of the cortex and hippocampus of mutant mice.

Results: Using ENU-directed mutagenesis, a new mutant mouse strain with epileptiform activity was generated. Through the modeling of audiogenic seizures (utilizing Krushinsky's method [4]), mice with a predisposition to epileptic seizures were identified, resulting in the development of a novel strain of mice named Socrates. Behavioral tests revealed that individuals of the Soc line exhibited reduced motor activity and learning ability, alongside increased orientation-exploratory activity, correlating with their heightened excitability. The development of a SNP panel using molecular genetic methods enabled the mapping of DNA regions to identify target genes involved in epilepsy formation. Subsequent transcriptomic analysis of these target genes demonstrated changes in their expression. Investigation into the expression patterns of genes encoding protein kinases, GABA and glutamate receptors, and membrane ion pathways revealed a decrease in expression, contributing to impaired neurotransmission in the cerebral cortex of Soc mice, ultimately being one of the factors behind the emergence of epileptic activity. Examination of the cytoarchitecture of the cortex and hippocampus revealed an increase in the number of interneurons in the cortex and an increase in the number of astrocytes in the hippocampus, indicating an imbalance in the excitation/inhibition balance in mutant individuals of the Soc line, which is also a contributing factor to the development of epileptiform states.

Conclusions: This study not only succeeded in creating and characterizing a new line of mice with an increased predisposition to epileptiform activity but also identified molecular genetic, morphophysiological, and cytoarchitectural changes leading to an altered balance between excitation and inhibition in the neuronal network of mutant mice.

Funding: This research was funded by the Ministry of Science and Higher Education of the Russian Federation (project FSWR-2023-0029).

Список литературы/References

1. Beghi E. The Epidemiology of Epilepsy. *Neuroepidemiology*. 2020;54(2):185-191
2. Patel D.C. et al. Neuron – glia interactions in the pathophysiology of epilepsy. *Nat Rev Neurosci*. 2019;20(5):282-297
3. Scheffer I.E. et al. Genetics of the epilepsies: Genetic twists in the channels and other tales. *Epilepsia*. 2010;51(S1):33-36
4. Vinogradova L.V. Neurophysiological mechanisms of reflex audiogenic epilepsy. Abstract of the dissertation of a doctor of biological sciences. Moscow, 2014

doi 10.18699/bgrs2024-7.2-03

Генетические проблемы суверенитета в биомедицинских исследованиях на лабораторных животных

Мошкин М.П.
Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия
mmp@bionet.nsc.ru

Медицинское импортозамещение предполагает формирование непрерывного цикла науки и производства, обеспечивающего продвижение в практику здравоохранения либо новых, либо репозиционированных лекарств. Обязательным звеном этого цикла является тестирование эффективности и безопасности средств профилактики и лечения болезней на лабораторных животных, среди которых первое место по масштабам использования занимают генетические линии мышей. И востребованность мышей фундаментальной наукой и фармакологией непрерывно растет, о чем свидетельствует позитивная динамика финансирования «мышиного животноводства». Тем не менее следует отметить, что использование животных в качестве тест-объектов критикуют не только озабоченные биоэтикой обыватели, но и профессиональные ученые. Последние обращают внимание на трудности получения воспроизводимых результатов, обусловленных фенотипической изменчивостью, присущей всем живым организмам.

С позиций современных требований тест-объекты доклинических испытаний должны: а) быть стандартизированными по своим феногенетическим характеристикам, б) быть свободными от возбудителей болезней (specific pathogen free – SPF), в) адекватно моделировать как массовые, так и орфанные патологии. Только при соблюдении этих требований возможно получение надежных результатов, в том числе и объединение результатов, полученных разными центрами неклинических испытаний. Блокирование контактов с мировыми депозитариями, поддерживающими разведение лабораторных животных в России, ставит вопрос об обеспечении функционирования российских питомников с сохранением высоких стандартов модельных организмов, опираясь на собственные технологические возможности. Итак, по порядку проблем:

Генетическое соответствие как основа поддержания феногенетического стандарта тест-объектов.

Спонтанные мутации и дрейф генов предопределяют неизбежные феногенетические изменения лабораторных линий животных. Несмотря на то что генетический дрейф имеет значение только в аутбредных популяциях, а частота спонтанных мутацией относится к крайне редким событиям, для гарантированного соответствия исходному генотипу племенные ядра меняют каждые 10 циклов размножения. Анализ отчетов биоресурсных коллекций (БРК), который проводился до 2021 г. в рамках одноименной программы ФАНО, показал, что в российских питомниках используют четыре подхода к поддержанию племенных ядер:

1. Практически бесконтрольное разведение с выбраковкой больных животных;

2. Разведение с выбраковкой особей с явным фенотипическим несоответствием разводимой линии и контролем совместимости кожных лоскутов;

3. Разведение и выбраковка при фенотипическом несоответствии, дополненные регулярной заменой племенных ядер из сертифицированных депозитариев;

4. Разведение и выбраковка при фенотипическом и генетическом несоответствии с регулярной заменой племенных ядер из собственного криоархива.

Как показал анализ отчетов биоресурсных коллекций (БРК), который проводился до 2021 г. в рамках одноименной программы ФАНО, только Центр генетических ресурсов (ЦГР) лабораторных животных ИЦиГ СО РАН обладает полным набором компетенций для суверенного обеспечения биомедицинских исследований модельными организмами надлежащего качества. В частности:

- наличие в ЦГР технологии криоконсервации позволяет преодолевать неизбежные последствия генетического дрейфа, выявляемого методом микросателлитного анализа, и поддерживать генетическое качество животных за счет обновления племенных ядер из криоархива;
- мониторинг здоровья в ЦГР в соответствии с европейскими требованиями (протокол FELASA) и использование вспомогательных репродуктивных технологий, включая экстракорпоральное оплодотворение, надежно обеспечивает поддержание SPF статуса;
- получение организмов, воспроизводящих патологии человека, обеспечено как собственным коллекционным фондом ЦГР – самым большим в РФ, так и наличием технологической платформы, обеспечивающей полный цикл создания требуемых моделей патологий методами геномного редактирования (CRISPR/Cas 9), селекции, хирургических вмешательств, фармакологии, манипуляциями с диетой и др.;
- верификация моделей патологий и оценка эффективности лекарственных и иных медицинских средств осуществляется широким набором методов прижизненного высокотехнологического фенотипирования, включая сверхвысокопольную магнитно-резонансную томографию;
- для выполнения сертифицированных протоколов определения безопасности новых лекарственных средств профилактики и лечения болезней на базе ЦГР создан Центр неклинических испытаний, аттестованный Росаккредитацией на соответствие требованиям надлежащей лабораторной практики (GLP).

Устойчивое функционирование ЦГР в условиях санкционного давления обеспечивается высококвалифицированными кадрами ЦГР и работающими на его базе научными сотрудниками ИЦиГ СО РАН. Таким образом, в современных реалиях ЦГР ИЦиГ СО РАН может стать центральной инфраструктурой, обеспечивающей поддержание, создание и распространение модельных организмов надлежащего качества.

Финансирование: Исследование выполнено при поддержке гранта РНФ № 23-14-00179.

Genetic challenges to sovereignty in biomedical research on animal subjects

Moshkin M.P.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
mmp@bionet.nsc.ru

Medical import substitution implies forming an integration cycle between research and industry that provides continuous adoption of new drugs into medical practice. The essential part of such an integration cycle is animal testing that ensures the effectiveness and safety of pharmaceutical agents and treatment strategies. Mice stains play a primary role in biomedical animal research, both in terms of genetic sophistication and in terms of sheer scale; the demand on genetic mice strains both in fundamental and applied research constantly grows as evidenced by the influx of investment into mouse breeding. However, it must be acknowledged that animal testing has its downsides. Apart from ethical concerns voiced by the general public, there is a professional critique coming from researchers themselves. The latter point out reproducibility problems that are inherent to animal testing because all living organisms have phenotypic variance.

By modern standards, clinical trial test subjects must: (a) have standard phenogenetic traits, (b) be specific pathogen free (SPF) and (c) adequately model mass and orphan pathologies. Only when these conditions are met, it is possible to produce reliable data and, importantly, to join the data acquired by different nonclinical trial institutions. Severing of the contacts with international depositaries, that have previously greatly facilitated laboratory animal breeding in Russia, poses a challenge to maintain steady work of vivaria and to keep the high standards of animal models, relying only on our own technological capabilities. In this light, one of the cardinal problems is the maintenance of animal test subjects' genetic coherence, which ensures a high phenogenetic standard of test subjects.

The phenogenetic discrepancy inside laboratory animal strains arises from spontaneous mutations and genetic drift. Despite the fact that genetic drift is significant only for outbred populations and the fact that non-neutral spontaneous mutations are extremely rare events, to guarantee the maintenance of the original genotype, the stock nucleus must be changed every 10 breeding cycles. Analysis of bioresource collections reports that was conducted until 2021 under the eponymous programme of Federal Agency of Scientific Organizations has shown that Russian vivaria practice 4 different approaches to maintaining stock nucleus:

1. Virtually uncontrolled breeding, only sick animals are removed;
2. Breeding with removal of animals with obvious phenotypic aberrations and control of skin graft compatibility;
3. Breeding with removal of phenotypic mismatches, augmented by regular replacement of stock nuclei from certified depositaries;
4. Breeding and removal of phenotypic and genotypic mismatches, with regular replacement of stock nuclei from the local cryo archive.

The reports from the above mentioned "bioresource collections" project indicate that only the Center for Genetic Resources (CGR) for laboratory animals of ICG SB RAS possesses the full spectrum of qualifications for independently supplying biomedical research with high quality model animals. In particular:

- The CGR has cryoconservation technology that allows to counter otherwise inevitable genetic drift. This is done by assessing the drift via DNA microsatellite analysis and resupplying the stock nucleus from cryo archive as needed;
- The CGR has animal health monitoring compliant with European standards (FELASA protocol) and utilizes assisted reproductive technologies, including in-vitro fertilization, which help to maintain the SPF animal status;
- The CGR can create animal models of human pathologies, as it has an extensive strain collection – largest in Russian Federation, as well as the full stack of technologies for genetic modification — genome editing methods (CRISPR/Cas 9), breeding and selection methods, surgical equipment and personnel, pharmacological administration methods, diet manipulation ability, etc;
- The CGR can perform verification of animal models of human pathology and assessment of pharmaceutical agents and therapies, due to a wide range of high-tech phenotyping methods, including super-high field magnetic resonance imaging;
- The CGR has a nonclinical trial center, which is accredited by Rosaccreditation (Russian accreditation) government agency for a good laboratory practices (GLP) compliance. This allows it to assess novel drug safety using certified protocols.

Stable functioning of the Center for Genetic Resources under economic sanctions regime is maintained by scientists of ICG SB RAS and other highly trained employees of the center. Thus, in today's challenging reality, the Center for Genetic Resources ICG SB RAS can become the central piece of biomedical infrastructure that ensures creation, maintenance, and distribution of high-quality animal models.

Funding: The study is supported by the Russian Science Foundation (grant No. 23-14-00179).

doi 10.18699/bgrs2024-7.2-04

Мыши с генетическим нокаутом IL-6 в модели STZ-индуцированного диабета

Пуртова С.К.*, Губернаторова Е.О., Горшкова Е.А.

Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия
**skpurtova@mail.ru*

Ключевые слова: диабет; интерлейкин 6; генетический нокаут; стрептозоцин; системное воспаление

Мотивация и цель: Интерлейкин 6 (IL-6) – ключевой провоспалительный цитокин, участвующий в патогенезе широкого спектра заболеваний, однако в последнее время появляется все больше данных, свидетельствующих о его вкладе в метаболические процессы и регенерацию тканей [1]. Чрезмерная и продолжительная продукция IL-6 может способствовать развитию патологии: так, известно, что хроническое воздействие IL-6 на периферические ткани и гепатоциты способствует развитию инсулинорезистентности, и, как следствие, диабета [2]. Роль IL-6 в развитии диабета и осложнений, ассоциированных с ним, не до конца изучена [3], поэтому целью нашего исследования являлась оценка тяжести патологии, в частности диабетической нефропатии, на фоне полного генетического нокаута IL-6.

Методы и алгоритмы: С целью оценки вклада IL-6 в развитие диабета у мышей дикого типа (WT) и мышей с генетическим нокаутом IL-6 (IL-6 KO) [4] был индуцирован диабет путем внутрибрюшинного введения стрептозоцина (STZ) ежедневно в течение пяти дней. Контрольным группам мышей вводили равный объем PBS. У животных в течение месяца измеряли вес, объем потребляемой воды и концентрацию глюкозы в крови. Также собирали мочу и сыворотку крови для дальнейшего биохимического анализа. На 30-й день эксперимента была проведена оценка инфильтрации иммунных клеток и уровня цитокинов в почках. Кроме того, проведен цитофлуориметрический анализ клеток селезенки и иммунных клеток, выделенных из почек.

Результаты: У мышей обоих генотипов, получавших STZ, развивались характерные симптомы диабета: гипергликемия, полиурия, полидипсия и глюкозурия. На четвертой неделе уровень глюкозы в крови и моче у мышей IL-6 KO значительно превышал показания, наблюдаемые у мышей WT. Также у них была более выражена потеря веса; таким образом, мыши IL-6 KO демонстрировали более тяжелое течение STZ-индуцированного диабета, чем мыши дикого типа. В конце эксперимента оценивали количество иммунных клеток в почках: у IL-6 KO мышей инфильтрация была значимо выше. Кроме того, в почках IL-6 KO наблюдалась тенденция к увеличению экспрессии хемокина CXCL9, что коррелировало с усиленной инфильтрацией лимфоцитов. Цитофлуориметрический анализ клеток, выделенных из селезенок, показал увеличение доли Т-регуляторных клеток и нейтрофилов у мышей с дефицитом IL-6 при индукции диабета, что также свидетельствовало о более тяжелом системном воспалении, ассоциированном с диабетом.

Выводы: Мыши с нокаутом по IL-6 во всем организме развивают более тяжелые симптомы диабета, а также более выраженную инфильтрацию селезенок и почек иммунными клетками, что может свидетельствовать о защитной роли этого цитокина в модели STZ-индуцированного диабета.

Финансирование: Исследование поддержано грантом РНФ № 19-75-30032.

IL-6 knockout mice in STZ-induced diabetes model

Purtova S.K.*, Gubernatorova E.O., Gorshkova E.A.

Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
** skpurtova@mail.ru*

Key words: diabetes; interleukin-6; genetic knockout; streptozotocin; systemic inflammation

Motivation and Aim: Interleukin 6 (IL-6) is a key proinflammatory cytokine, which also contributes to metabolic processes and tissue regeneration [1]. At the same time, excessive and prolonged production of IL-6 may contribute to the development of pathology. It is known that chronic exposure to IL-6 of peripheral tissues and hepatocytes leads to the inflammation, which induces insulin resistance and overt diabetes [2]. The role of IL-6 in the development of diabetes and diabetes-associated complications is not fully understood [3]. Therefore, the aim of our study was to assess the severity of diabetes and diabetic nephropathy in IL-6-deficient mice (IL-6 KO).
Methods and Algorithms: To study the role of IL-6 in diabetes wild-type (WT) and IL-6 KO [4] mice were injected intraperitoneally with streptozotocin (STZ) for five days. The control groups of both genotypes were administered an equal volume of PBS. Mice were monitored for weight changes, water intake, and blood glucose levels for a period of one month. Urine and serum samples were also collected for the analysis. An assessment of the immune cell infiltration in the kidneys and a cytometric analysis of splenocytes were carried out on the 30th day of the experiment.
Results: Mice of both genotypes treated with STZ developed the characteristic symptoms of diabetes: hyperglycemia, polyuria, polydipsia and glycosuria. At week four after diabetes induction, blood and urine glucose levels in IL-6 KO mice were significantly higher than those observed in WT mice. Weight loss was more pronounced in IL-6 KO. Thus, IL-6 KO mice exhibited a more severe course of STZ-induced diabetes than WT mice. At the end of the experiment, the number of immune cells infiltrating the kidneys was assessed, the infiltration was significantly higher in IL-6 KO mice. In addition, IL-6 KO kidneys tended to have increased levels of CXCL9, which correlated with lymphocyte infiltration. FACS analysis of splenocytes showed an increase in the proportion of regulatory T cells and neutrophils in IL-6-deficient mice upon induction of diabetes, which also indicated more severe systemic inflammation associated with diabetes.
Conclusion: IL-6 deficient mice develop more severe symptoms of diabetes, as well as more pronounced immune cell infiltration of the kidneys related, which may indicate a protective role for this cytokine in the STZ-induced diabetes model.
Funding: The study was supported by RSF grant No. 19-75-30032.

Список литературы/References

1. Scheller J., Chalaris A., Schmidt-Arras D., Rose-John S. The pro- and anti-inflammatory properties of the cytokine interleukin-6. *Biochim Biophy Acta.* 2011;1813(5):878-888
2. Rehman K., Akash M.S.H., Liaqat A. et al. Role of Interleukin-6 in development of insulin resistance and type 2 diabetes mellitus. *Crit Rev Eukaryot Gene Expr*. 2017;27(3):229-236
3. Su H., Lei C.T., Zhang C. Interleukin-6 signaling pathway and its role in kidney disease: An Update. *Front Immunol*. 2017;8:405
4. Gubernatorova E.O., Gorshkova E.A. Non-redundant Functions of IL-6 Produced by Macrophages and Dendritic Cells in Allergic Airway Inflammation. *Front Immunol.* 2018;9:2718

doi 10.18699/bgrs2024-7.2-05

Междисциплинарные подходы в исследованиях ПТСР: от экспериментального моделирования к внедрению в клинику

Цейликман В.[1]*, Цейликман О.[1, 2], Карпенко М.[3], Майстренко В.[3], Шатилов В.[2], Цейликман Д.[2], Шамшурин М.[2], Грабчук М.[2], Бурлаков Е.[2], Бричагина А.[5], Жуков М.[2], Липатов И.[2], Букша И.[2], Аристов М.[2], Эпиташвили А.[2], Шонина А.[2], Колесникова А.[2], Ермолнеко А.[4]

[1] *Южно-Уральский государственный университет, Челябинск, Россия*

[2] *Челябинский государственный университет, Челябинск Россия*

[3] *Институт экспериментальной медицины, Санкт-Петербург, Россия*

[4] *Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*

[5] *Московский физико-технический институт, Долгопрудный, Россия*

* *tceilikmanve@susu.ru*

Ключевые слова: ПТСР; моноаминооксидазы; моноамины; нейротрансмиттеры; генные сети

Мотивация и цель: ПТСР развивается как отдаленное последствие пережитой психологической травмы. До сих пор до 40 % больных ПТСР проявляют устойчивость к действию фармакологических препаратов [1]. Наличие подобных барьеров связано в том числе с недостаточной изученностью патогенеза ПТСР. Целью данной работы является комплексное изучение механизмов ПТСР на основе междисциплинарного подхода, сочетающего экспериментальные исследования с методическим арсеналом биоинформатики.

Методы и алгоритмы: У крыс Вистар ПТСР моделировался запахами хищника. По значению индекса тревожности (ИТ) животные подразделялись на подверженных (ПТСР+) и устойчивых (ПТСР–) фенотипы. Различия по ПТСР+ и ПТСР– фенотипам анализировались по уровню экспрессии генов, нейрохимическим, эндокринологическим, метаболическим и гистологическим показателям. Для анализа микроРНК, специфичных к экспрессии МАО-А и МАО-В, был написан программный код на языке Python, сравнивающий Integration Rate микроРНК и совершающий их кластеризацию.

Результаты: У ПТСР+ крыс по сравнению с ПТСР– крысами наблюдалось выраженное снижение концентрации кортикостерона в крови, уровней дофамина и серотонина с одновременным увеличением в гиппокампе экспрессии генов МАО-А, катехоламино-О-метилтрансферазы (КОМТ) при снижении экспрессии гена BDNF. В печени ПТСР+ крыс повысились активность и экспрессия фермента 11β гидроксистероиддегидрогеназы1 (11βГСДГ-1), вовлеченного в метаболизм глюкокортикоидов. Отмечена положительная корреляция между активностью 11βГСДГ-1 и значением ИТ ($r = +0.95$; $p<0.05$). В печени и миокарде ПТСР+ крыс отмечено наличие окислительного стресса, митохондриальной дисфункции и гистологических признаков повреждения органов дистрофического характера. Кроме того, для ПТСР+ крыс отмечены аномалии ЭКГ, признаки дислипопротеинемии и гипогликемии. Экспериментальная коррекция ПТСР осуществлялась ресвератролом в дозах (100, 50, 20, 10 мг/кг), ингибиторами обратного захвата

серотнина – пароксетином (10 мг/кг), флуоксетином (10 мг/кг), сертралином (10 мг/кг), циталопрамом (20 мг/кг). Полученные результаты показывают ощутимое преимущество ресвератрола в коррекции ПТСР по сравнению с препаратами первой линии. Высокая эффективность ресвератрола объясняется его способностью корригировать вызванные ПТСР сбои в генных сетях через активацию деацетилазной активности сиртуина 1 (SIRT1). Дополнительно ресвератрол может корригировать ПТСР путем ингибирования активности 11βГСДГ-1. Но при этом выявлен фенотип ПТСР+ крыс, устойчивых к ресвератролу. В отличие от сенсетивных к терапии крыс для них характерен повышенный уровень активности/экспрессии МАО-А, но не МАО-В. Полученные результаты согласуются с данными о способности SIRT1 парадоксально усиливать ПТСР через активацию SIRT1/NHLH2/MAO-A путь по причине недостаточного негативного контроля со стороны микроРНК [2]. Анализ кластеризации микроРНК для МАО-А и МАО-В, показал отсутствие среди них совпадений, что для двух случайных генов составляет вероятность всего 3.2 %. Соответственно в генных сетях ПТСР сбои могут реализоваться на уровне регуляции экспрессии этих ферментов.

Выводы: Полученные результаты свидетельствуют о наличии отчетливых межфенотипических различий между ПТСР+ и ПТСР– крысами не только на поведенческом, но и на нейрохимическом, нейроэндокринном, метаболическом и молекулярном уровнях. Благодаря междисциплинарным подходам удалось установить вклад нарушений генных сетей в патогенез ПТСР и обосновать наилучшую эффективность ресвератрола как компонента ассоциативной генной сети в коррекции ПТСР. Также была создана диагностическая панель с МАО-А/МАО-В микроРНК, которой можно эффективно воспользоваться для персонализированной терапии ПТСР.

Финансирование: Исследование поддержано региональным грантом РНФ_Челябинская область № 23-15-20040.

Interdisciplinary approaches in PTSD research: from bench to bedside

Tseilikman V.[1]*, Tseilikman O.[1, 2], Karpenko M.[3], Maistrenko V.[3], Shatilov V.[2], Tseilikman D.[2], Shamshurin M.[2], Grabchuk M.[2], Burlakov E.[2], Brichagina A.[5], Zhukov M.[2], Lipatov I.[2], Buksha I.[2], Aristov M.[2], Epitashvili A.[2], Shonina A.[2], Kolesnikova A.[2], Ermolneko A.[4]

[1] *South Ural State University, Chelyabinsk, Russia*

[2] *Chelyabinsk State University, Chelyabinsk, Russia*

[3] *Institute of Experimental Medicine, St. Petersburg, Russia*

[4] *St. Petersburg State University, St. Petersburg, Russia*

[5] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

* *tceilikmanve@susu.ru*

Ключевые слова: ПТСР; моноаминооксидазы; моноамины; нейротрансмиттеры; генные сети

Motivation and Aim: PTSD develops as a long-term consequence of experienced psychological trauma. Until now, up to 40 % of PTSD patients show resistance to the

effects of pharmacological drugs [1]. The presence of such barriers is associated, among other things, with insufficient knowledge of the pathogenesis of PTSD. The purpose of this work is a comprehensive study of the mechanisms of PTSD based on an interdisciplinary approach that combines experimental research with the methodological arsenal of bioinformatics.

Methods and Algorithms: In Wistar rats, PTSD was performed by predator odors. According to the Anxiety Index (AI) value, animals were divided into susceptible (PTSD+) and resistant (PTSD–) phenotypes. Differences in PTSD+ and PTSD– phenotypes were analyzed by the level of gene expression, neurochemical, endocrinological, metabolic and histological parameters. To analyze microRNAs specific to the expression of MAO-A and MAO-B, a program code was written in Python that compares the Integration Rate of microRNAs and performs their clustering.

Results: It has been revealed decrease of the plasma corticosterone concentration as well as decrease the dopamine and serotonin levels in the hippocampus, what were associated with an increase in the expression of MAO-A and catecholamine-O-methyltransferase (COMT) genes in PTSD+ rats in compared with PTSD- rats. Simultaneously a decrease in the expression of the BDNF gene also was indicated. In the liver of PTSD+ rats, the activity and expression of 11β hydroxysteroid dehydrogenase 1 (11β HSD-1), a main enzyme involving in glucocorticoid metabolism was increased. There was a positive correlation between the activity of 11β HSD-1 and the AI value ($r = +0.95$; $p < 0.05$). In the liver and myocardium of PTSD+ rats, the presence of oxidative stress, mitochondrial dysfunction and histological signs of dystrophic organ damage were noted. In addition, ECG abnormalities, signs of dyslipoproteinemia and hypoglycemia were indicated PTSD+ rats.

Experimental correction of PTSD was carried out with resveratrol in doses (100, 50, 20, 10 mg/kg), serotnine reuptake inhibitors – paroxetine (10 mg/kg), fluoxetine (10 mg/kg), sertraline (10 mg/kg), citalopram (20 mg/kg). The results obtained show a significant advantage of resveratrol in the correction of PTSD compared to first-line drugs. The high effectiveness of resveratrol is explained by its ability to correct PTSD-induced disruptions in gene networks through activation of the deacetylase activity of sirtuin 1 (SIRT1). Additionally, resveratrol can correct PTSD by inhibiting 11βHSD-1 activity. However, a phenotype of PTSD+ rats resistant to resveratrol was revealed. Unlike rats sensitive to therapy, they are characterized by an increased level of activity/expression of MAO-A, but not MAO-B. The results obtained here are consistent with data on the ability of SIRT1 to paradoxically enhance PTSD manifestation through activation of the SIRT1/NHLH2/MAO-A pathway due to insufficient negative control by microRNAs [2]. Analysis of microRNA clustering for MAO-A and MAO-B showed the absence of matches among them, which for two random genes is only 3.2 % probability. Accordingly, in PTSD gene networks, disruptions can occur at the level of regulation of the expression of these enzymes.

Conclusion: The results obtained here indicate the presence of distinct differences between PTSD+ and PTSD– rats not only at the behavioral but also at the neurochemical, neuro-endocrine, metabolic and molecular levels. Thanks to interdisciplinary approaches, it was possible to establish the contribution of gene network disorders in the pathogenesis of PTSD and substantiate the best effectiveness of resveratrol as a component of an associative gene network in the correction of PTSD. A diagnostic panel with MAO-A/MAO-B microRNA was also created, which can be effectively used for personalized therapy for PTSD.

Funding: The study is supported regional grant of the Russian Science Foundation_Chelyabinsk region No. 23-15-20040.

Список литературы/References

1. Williams T., Phillips N.J., Stein D.J., Ipser J.C. Pharmacotherapy for post traumatic stress disorder (PTSD). *Cochrane Database Syst Rev*. 2022;3(3):CD002795. doi 10.1002/14651858.CD002795.pub3
2. Li W., Guo B., Tao K., Li F., Liu Z., Yao H., Feng D., Liu X. Inhibition of SIRT1 in hippocampal CA1 ameliorates PTSD-like behaviors in mice by protections of neuronal plasticity and serotonin homeostasis via NHLH2/MAO-A pathway. *Biochem Biophys Res Commun*. 2019;518(2):344-350. doi 10.1016/j.bbrc.2019.08.060

doi 10.18699/bgrs2024-7.2-06

Db/db mice as diabetes genetic model, effect of trehalose, comparative study of liver, heart and brain

Korolenko T.A.[1]*, Bgatova N.P.[2], Zavjalov E.L.[3], Korolenko E.[4], Goncharova N.V.[1], Pupyshev A.B.[1], Tenditnik M.V.[1], Johnston T.P.[5]

[1] *Scientific Research Institute of Neuroscience and Medicine, Novosibirsk, Russia*

[2] *Laboratory of Ultrastructural Research, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[4] *University Canada West, Vancouver, Canada*

[5] *Division of Pharmacology and Pharmaceutical Sciences, School of Pharmacy, University of Missouri-Kansas City, Kansas City, Missouri, USA*

* *box2023tatiana@mail.ru*

Key words: diabetes model; db/db mice; autophagy; trehalose

Type 2 diabetes is a prevalent metabolic disorder, with steady increasing of numbers of cases [1, 2]. Several experimental models have been proposed for studies on Type 2 diabetes pathogenesis, the db/db mice are known genetic animal model of diabetes Type 2 [1–3] used for choice in experimental therapy. This phenotype includes severe obesity, hyperphagia, polydipsia, and polyuria, connected with a spontaneous mutation of leptin receptor [1, 2].

It was shown, that model of Type 2 diabetes in db/db mice was characterized also by intracellular lipid storage in hepatocytes and cardiomyocytes and some signs of inflammation (increased expression of TNF-α in the spleen, underexpression of IL-10 in the liver and spleen [3]. These changes can play role in development of early pathological changes of atherosclerosis (heart) and brain, the main complications of diabetes T2 in human.

The Aim of this study was evaluating the effect of autophagy inducer trehalose on liver, heart and brain ultrastructure, with concentration of attention at autophagy and intracellular lipids storage in cells.

Methods: 38 male db/db mice, aged 3 months at time 0 of acclimation (SPF vivarium of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk) and C57BL/6 mice (as a control) were used.

Experimental Design: The mice were subdivided into four groups (6–8 animals each):

1) "C57BL/6 mice", i. e., C57BL/6 mice drinking water ad libitum during the whole experiment (24 days);

2) "trehalose-treated C57BL/6 mice", i. e., C57BL/6 mice drinking a 2 % solution of trehalose (Trehalose dihydrate, Tokyo Chemical Industry, Japan) instead of water;

3) "db/db mice", i. e., db/db mice drinking water ad libitum during the whole experiment; and

4) "Trehalose-treated db/db mice", i. e., db/db mice drinking the 2 % trehalose aqueous solution. On the day of euthanasia (at day 25), the animals in each group were killed by decapitation.

Biochemical Assays: Murine blood was collected after decapitation, and serum was obtained by centrifuge Eppendorf 5415R Eppendorf AG, Hamburg, Germany) at

3000× g for 20 min. Fasting blood glucose, glycosylated hemoglobin HA1, and liver function (ALT activity) were assayed during the experiment [3].

Transmission Electron Microscopy: Myocardium, liver and brain (prefrontal cortex) samples for electron microscopy were fixed in 4 % paraformaldehyde in the Hanks medium and a 1 % OsO4 solution (Sigma, St. Louis, MO, USA) in phosphate buffer (pH 7.4) for 1 h, dehydrated in ethanol of ascending concentrations, and embedded in Epon (Serva) (3) and analyzed under a JEM 1400 electron microscope (JEOL Ltd., Tokyo, Japan) (Multiple-Access Centre for Microscopy of Biological Subjects, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia).

Morphometric Electron-Microscopic Analysis: To determine volume density of autophagic structures in hepatocytes and cardiomyocytes, 30 cells in each group were randomly selected. Volume densities of autophagosomes, autolysosomes, and lysosomes were calculated using the ImageJ software (National Institutes of Health, Bethesda, MD, USA). Autophagic structures were identified according to the guidelines for monitoring autophagy.

Morphometric Analysis of Lipid Inclusions: Quantitative data were obtained using the iTEM software (Olympus, Tokyo, Japan). Results are presented as the mean ± SEM, with $p < 0.05$ were regarded as statistically significant.

Results: Trehalose treatment of C57L/6 mice significantly reduced blood glucose concentration and glycated hemoglobin. In comparison with C57BL/6 mice, blood glucose concentration was significantly ($p < 0.001$) higher in db/db mice, as was glycated hemoglobin ($p < 0.001$); trehalose treatment significantly ($p < 0.01$) reduced both parameters. Treatment of db/db mice by trehalose was followed by increased autophagy induction in the heart and liver.

Compared to the control (C57BL/6 mice), the weight of db/db mice (drinking plain water) was greater because of obesity ($p < 0.001$), whereas trehalose consumption decreased this parameter in comparison to untreated animals ($p < 0.01$). Heart weight decreased by trehalose consumption in db/db mice as compared to untreated mice ($p < 0.05$). Liver weight was greater both in untreated ($p < 0.001$) and trehalose-treated db/db mice ($p < 0.01$) vs respective controls. The consumption of trehalose with drinking water decreased the body weight of db/db mice compared to untreated db/db mice ($p < 0.05$). Spleen weight was ~2-fold greater in untreated db/db mice than in untreated C57BL/6 mice ($p < 0.001$), possibly attributable to enhanced functioning of the spleen in Type 2 diabetes as a consequence of liver steatosis.

In hepatocytes of db/db mice (compared to the control C57Bl/6 mice) there was a higher content of large lipid inclusions, a lower amount of glycogen, and a larger number of convoluted annular mitochondria. In the cytoplasm of hepatocytes from trehalose-treated db/db mice, we observed autophagosomes with cytoplasm fragments, and autolysosomes with visualized lipid inclusions.

In myocardium there was a moderate deficit of the sarcoplasm owing to lytic alterations. A distinctive feature of the cardiomyocyte structure in db/db mice was the presence of a large number of lipids droplets in their sarcoplasm, which were overall evenly distributed within a cell. Increased lipid droplets were observed also in some neurons of brain (perifrontal zone) of db/db mice.

Conclusion: The autophagy activation by trehalose revealed several positive effects on the heart and liver of db/db mice possibly via increased lipophagy (uptake of lipid droplets). Therefore, lipophagy activation seems to be a promising approach to prevent complications in diabetes.

Funding: The study is supported by budget grant (No. 122042700001-9).

References

1. Zhang Y., Pan X.F., Chen J. et al. Combined lifestyle factors and risk of incident type 2 diabetes and prognosis among individuals with type 2 diabetes: A systematic review and meta-analysis of prospective cohort studies. *Diabetologia*. 2020;63:21-33
2. Banach M., Surma S., Reiner Z. et al. Personalized management of dyslipidemias in patients with diabetes – it is time for a new approach. *Cardiovasc Diabetol*. 2022;21(1):263. doi 10.1186/s12933-022-01684-5
3. Korolenko T.A., Ovsyukova M.V., Bgatova N.P. et al. Trehalose Activates Hepatic and Myocardial Autophagy and Has Anti-Inflammatory Effects in *db/db* Diabetic Mice. *Life*. 2022;12(3):442. doi 10.3390/life12030442

doi 10.18699/bgrs2024-7.2-07

Ceramide affects intestinal epithelial barrier integrity via destabilization of actin cytoskeleton in chronic colitis

Medvedeva S.[1], Boldyreva L.[2]*, Achasova K.[1]*, Ogienko A.[1], Kozhevnikova E.[1, 3]**

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[2] *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

[3] *Novosibirsk State Agrarian University, Novosibirsk, Russia*

* *Equal contribution*

** *kozhevnikova@mcb.nsc.ru*

Key words: chronic colitis; intestinal barrier; filamentous actin; lipid metabolism

Motivation and Aim: Intestinal barrier dysfunction, also known as “leaky gut”, is characteristic of gastrointestinal diseases, especially inflammatory bowel diseases (IBD), celiac disease, enteric infections, irritable bowel syndrome (IBS), and others. Despite the extensive research in this field, there are only a few drugs available for the treatment of IBD in clinical practice. Although numerous works have investigated the mechanisms underlying the regulation of gut barrier function using models of acute inflammation, there is still a lack of understanding about how the barrier function of the intestine is compromised during the development of low-grade inflammation.

Methods: Metabolomic and transcriptomic analyses were combined with three alternative in vivo models of chronic colitis in mice to unravel common metabolic features of inflammation. Confocal microscopy was used to localize filamentous actin and tight- and adherence junctions’ proteins in the intestinal tissues. In vivo intestinal permeability was used in the functional test of barrier function.

Results: Here we show that phospholipid metabolism is impaired in transgenic Muc2 mouse model of chronic colitis, which is supported by the transcriptomic and metabolomic analysis of the descending large intestine. This data was supported in two alternative models of chronic colitis in mice in vivo: chemically-induced and associated with T-cells adoptive transfer colitis. We proposed that the upregulated ceramide metabolism might be responsible for the impairment of filamentous actin dynamics impairment, which in turn controls barrier function. The addition of ceramide and ceramide-synthase inhibitors affect actin polymerization, tight- and adherence junction proteins localization to the membrane, and intestinal barrier function in two independent mouse colitis models *in vivo*. This data agrees with current studies suggesting upregulation of ceramide in IBD patients.

Conclusion: Thus, our data provides a novel mechanism that underlies intestinal barrier disfunction upon chronic colitis.

Funding: The study is supported by funded by the Russian Science Foundation, grant number 20-74-10022-П.

doi 10.18699/bgrs2024-7.2-08

Disfunction of intercellular junctions in chronic colitis models

Medvedeva S.S.[1]*, Boldyreva L.V.[2], Achasova K.M.[1, 2], Ogienko A.A.[1], Kozhevnikova E.N.[1]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[2] *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

* *medvedeva@mcb.nsc.ru*

Key words: chronic colitis; F-actin; tight junctions; adherence junctions; leaky gut

Motivation and Aim: Increased permeability of the intestinal barrier, commonly known as "leaky gut syndrome," is a prevalent functional disorder observed in various gastrointestinal diseases, including inflammatory bowel diseases (IBD), celiac disease and bowel infections. The underlying cause of leaky gut syndrome is the disruption to the structure of tight junctions (TJ) and adherence junctions (AJ) within the intestinal epithelium. Numerous studies focusing on acute IBD suggest that an imbalance between pro- and anti-inflammatory factors leads to impaired distribution of TJ and AJ proteins [1–3]. Researches on F-actin reorganization during inflammation *in vivo* are non-systemic and contradictory, although microfilaments are essential for TJ and AJ formation [4, 5]. We aimed to establish a connection between F-actin stability and intercellular junctions in the context of leaky gut which can develop in various ways. Here, we describe F-actin, TJ and AJ structure and distribution in colons of three murine models of chronic colitis: *Muc2*-KO mice [6], DSS-induced chronic colitis [7] and $CD4^{+}CD45RB^{High}$ cell transfer model [8].

Methods and Algorithms: *Muc2*-KO mice were obtained by the rederivation of previously generated $Muc2^{tm1Avel}/Muc2^{tm1Avel}$ mice on C57BL/6 genetic background in SPF CD1 female mice and backcrossing to C57BL/6JNskrc. For DSS chronic inflammation, WT mice received three cycles of 2 % DSS/water. For cell transfer inflammation, athymic 12-week-old male Nude (Nu/j) mice were injected with sorted $CD4^{+}CD45RB^{High}$ cells obtained from Balb/c mice.

Distribution of AJ and TJ proteins in the descending colon was assessed with immunohistochemistry, using anti-β-catenin (# 610153, BD Transduction Laboratories), anti-Claudin 7 antibodies (# 37-4800, Invitrogen) and Alexa Fluor 568 Phalloidin (# A12380, ThermoFisher Scientific) for F-actin tracking.

To study F-actin interactions, WT mice were administered Latrunculin A (# 100-0562, Stemcell Technologies) or Jasplakinolide (# sc-202191A, Chemcruz) rectally. F-actin polymerization was evaluated by filming fluorescence recovery after photobleaching (FRAP) with CellMask Green Actin dye (# A57243, Invitrogen).

Results: Immunohistochemistry analysis revealed that modulators of F-actin stability Latrunculin A and Jasplakinolide affect not only microfilaments but Claudin 7 and β-catenin localization as well. FRAP timelapses showed decreased polymerization rate of actin filaments in enterocytes of *Muc2*-KO mice. All three models of chronic inflammation demonstrated disrupted structure of actin microfilaments, which anchor the TJ and AJ complexes [5]. Additionally, the localization of the TJ protein Claudin 7 and AJ protein β-catenin on the lateral membranes of enterocytes was also impaired in all three models.

Conclusion: Disruptions in F-actin structure, coupled with altered localization of Claudin 7 and β-catenin, appear to be common features across at least three mouse models of chronic colitis. These features are, potentially, universal characteristics of increased permeability in chronic colitis. Therefore, we suggest that the mechanism of leaky gut syndrome development is as follows: chronic inflammation affects F-actin structure and causes TJ and AJ instability.

Funding: The study is supported by the Russian Science Foundation (RSF) Grant (No. 20-74-10022-П).

References

1. Oshima T., Miwa H., Joh T. Changes in the expression of claudins in active ulcerative colitis. *J Gastroenterol Hepatol*. 2008;23:S146-S150
2. Luettig J. et al. Claudin-2 as a mediator of leaky gut barrier during intestinal inflammation. *Tissue Barriers*. 2015;3(1-2):e977176
3. Lechuga S., Ivanov A.I. Disruption of the epithelial barrier during intestinal inflammation: Quest for new molecules and mechanisms. *Biochim Biophys Acta Mol Cell Res*. 2017;1864(7):1183-1194
4. Ivanov A.I., Parkos C.A., Nusrat A. Cytoskeletal regulation of epithelial barrier function during inflammation. *Am J Pathol*. 2010;177(2):512-524
5. Belardi B. et al. A weak link with actin organizes tight junctions to control epithelial permeability. *Developmental Cell*. 2020;54(6):792-804.e7
6. Mizoguchi A. et al. Genetically engineered mouse models for studying inflammatory bowel disease. *J Pathol*. 2016;238(2):205-219
7. Randhawa P.K. et al. A review on chemical-induced inflammatory bowel disease models in rodents. *Korean J Physiol Pharmacol*. 2014;18(4):279-288
8. Steinbach E.C., Gipson G.R., Sheikh S.Z. Induction of murine intestinal inflammation by adoptive transfer of effector CD4+ CD45RBhigh T cells into immunodeficient mice. *J Vis Exp*. 2015;98:52533

doi 10.18699/bgrs2024-7.2-09

Functional characteristics of previously unstudied genes associated with congenital malformations of the nervous system

Mitina N.N.[1]*, Filat'eva A.E.[1], Anisimova P.E.[1], Kondakova E.V.[1], Tarabykin V.S.[2]

[1] *Research Institute of Neurosciences, Lobachevsky Nizhny Novgorod State University, Nizhny Novgorod, Russia*

[2] *Institute of Cell Biology and Neurobiology, Charite Clinic, Berlin, Germany*

**turyginanatasha@yandex.ru*

Key words: neurodevelopmental disorders; mental retardation; in utero electroporation; corticogenesis

Motivation and Aim: Hereditary diseases, including those associated with neuro-developmental disorders, are one of the key problems for the health care system.
For a large number of patients with undifferentiated forms of mental retardation, intellectual development delay, and autism, the molecular mechanisms underlying neurodevelopmental disorders are still unknown. Associated pathologies include developmental delay, spasticity, microcephaly, and primary epilepsy. Understanding the function of previously unstudied genes is essential for diagnosis, prognosis of the disease, and the quality of medical genetic counseling. In this project, we use modern technologies to create in vivo models and to obtain new data on molecular mechanisms leading to disorders of nervous system development.
Methods and Algorithms: To visualize the expression pattern at embryonic (E14.5, E16.5, E18.5) stages of development, the molecular biological method of RNA in situ hybridization (ISH) was used. RNA probes are pre-synthesized with cDNA of the target gene using gene-specific primers. Subsequently, the target mRNA of the gene is hybridized with RNA probes labeled with DIG-dUTP (digoxigenin). If the target mRNA is present in the sample, then the complementary nucleotide sequences of the sample and the mRNA form stable double-stranded bonds, which subsequently makes it possible to identify the location of tissue areas where the target gene is expressed.
To down-regulate the expression of studied genes specifically in cortical progenitors in vivo, we performe injections of plasmid constructs expressing shRNA targeting gene mRNA into the lateral ventricles of E13.5 mouse embryos followed by in utero electroporation for transfection. Concurrently, we co-electroporate these cells with a plasmid containing green fluorescent protein (GFP) cDNA to facilitate cell tracking. We then prepare histological slides of the electroporated cortex by isolating and fixing embryonic brain samples at days 15.5 and 17.5 of gestation and perform immunohistochemical analysis of control and KO tissues.
To analyze mitosis exit, we inject BrdU, a thymidine nucleoside analog, during IUE. BrdU is incorporated into the DNA of dividing cells at S phase of the mitotic cycle. Twenty-four hours post-electroporation, pregnant mice undergoes injection with a single dose of BrdU. Cortices are collected 2 days after IUE, fixed, and immunolabeled against EGFP, ki67, and BrdU. Quantification of both EGFP and BrdU-positive neurons relative to the total number of EGFP+ cells reveals how many neurons were born from electroporated APs and indicates whether the target gene regulates AP exit from the cell cycle. Quantification of the fraction of ki67-positive cells (a marker of undergoing

mitosis) should reveal any differences in the ability of progenitors to divide between control and KO tissues.
To study the effect of KO on cell death, we assess the expression of cleaved caspase-3: quantitative analysis of EGFP+ and caspase-3+ cells in the EGFP+ cell population reveals whether loss of a gene causes neuronal death during neurogenesis.
To study the effect of KO on migration, the distribution of migrating neurons in the layers of the developing cortex was assessed. For this purpose, sections of electroporated cortex collected on days 15.5 and 17.5 of gestation were used, as described above. Differences in the distribution pattern of GFP+ cells indicate whether KO affects the speed and direction of neuronal migration. Studying the morphology of GFP+ cells and assessing the ratio of polarized and multipolar cells show whether KO affects polarization processes in developing neurons
Results: We examined the expression pattern of two candidate genes in mouse brain sections using in situ hybridization (ISH). For one of them, ubiquitous expression was found at all developmental stages studied. For the second gene expression was found in the ventricular and subventricular zones, and weaker expression was detected in the cortical plate and in the intermedial zone.
In control electroporation experiments, approximately 20 % of all GFP-positive cells were found to be double-positive for ki67, indicating that a subset of neocortical cells incorporating the targeting constructs remained actively cycling 48 hours post-electroporation [1]. However, with shRNA electroporation, the proportion of cells persisting in the mitotic cycle doubled compared to control conditions. These observations suggest that studied gene plays a crucial role in regulating the cell cycle progression of forebrain neuronal stem cells. Downregulation of its expression results in prolonged cell cycle duration and reduced initiation of neuronal differentiation, potentially leading to a decreased number of neurons generated in the neocortex.
Funding: The study was supported by the Ministry of Science and Higher Education of the Russian Federation (No. FSWR-2023-0029).

References

1. Dharmadhikari A.V., Abad M.A., Khan S., Maroofian R. et al. RNA methyltransferase SPOUT1/CENP-32 links mitotic spindle organization with the neurodevelopmental disorder SpADMiSS. *medRxiv*. 2024. doi 10.1101/2024.01.09.23300329

doi 10.18699/bgrs2024-7.2-10

New avenues for treatment with trehalose of mice with pharmacological and transgenic models of Alzheimer's disease

Pupyshev A.B.*, Akopyan A.A., Tenditnik M.V., Ovsyukova M.V., Dubrovina N.I., Tikhonova M.A.

Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia
** pupyshevab@neuronm.ru*

Key words: Alzheimer's disease, trehalose, autophagy, amyloid-β 25-35, 5xFAD mice, hippocampus, LC3-II, IBA1, passive avoidance

Motivation and Aim: Recently, the treatment of animal models of Alzheimer's disease (AD) using autophagy activation has been actively developed. It can attenuate the accumulation of aberrant proteins, activate mitophagy, reduce oxidative stress and neuroinflammatory process [1, 2]. Positive results have been obtained for both inducers of mTOR-dependent autophagy (rapamycin and its analogs) [3, 4] and activators of mTOR-independent autophagy (lithium, curcumin, metformin) [5]. Induction of mTOR-independent autophagy, in contrast to mTOR-dependent autophagy, does not cause concomitant pleiotropic effects of suppression of biosynthetic processes, inhibition of cell growth and proliferation, and suppression of immunity [6]. The extent to which these effects are therapeutically beneficial or detrimental for the treatment of neurodegeneration is not known. Among mTOR-independent inducers, the disaccharide trehalose has attracted particular attention for its ability to reduce Aβ accumulation, produce antioxidant and anti-inflammatory effects, and restore hippocampal neurons and cognitive function [7–9]. Therapeutic effects have been identified for both pharmacological and transgenic models of the disease [9–11]. The predominant mode of treatment is the addition of trehalose in drinking (2% solution) [12, 13]. Nevertheless, it is of interest to find the optimal regimens for trehalose administration [7, 9] and reveal the efficacy of trehalose treatment depending on the age of mice [4, 14], since it is important to know the effectiveness of treatment at late stages of disease course, close to the translational transition.
Methods and Algorithms: A dose-dependent therapeutic effect of trehalose (2, 4, and 4% in the intermittent mode) was studied in a pharmacological model of AD induced by intracerebroventricular injection of aggregated amyloid-β fragment Aβ (25–35) [15]. To elucidate the efficacy of age-dependent trehalose treatment, mutant mice of 5xFAD strain with progressive course of the disease were used. Three groups of mice were studied: 1. Trehalose treatment (2 months) of 2-month-old mice. 2. Trehalose treatment (5 months) of 2-month-old mice. 3. Trehalose treatment (2 months) of 5-month-old mice. In mice of the 5xFAD strain, behavioral abnormalities have appeared at the age of 5–6 months. Behavioral testing was performed using open-field, plus-maze, and conditioned passive avoidance response tests. At the end of testing, brain material was collected for immunohistochemical analysis (IHC) using transcardiac perfusion of the brain with saline solution containing 4 % paraformaldehyde. In the frontal cortex and three hippocampal regions (CA1, CA3, and dentate gyrus) affected in AD, the expression

of autophagy marker LC3-II and neuroinflammatory marker IBA1, Aβ accumulation, and neuronal density (Nissl staining) were evaluated [15, 16].

Results: Pharmacomodeling of AD by administration of Aβ (25–35) caused a marked increase in the levels of the disease marker Aβ in the frontal cortex and all hippocampal areas studied, while all variants of two-week trehalose treatment sharply reduced the index ($p < 0.01$ for the CA1 hippocampal area), restoring it almost to baseline. In parallel, a significant activation of autophagy was noted, especially for 4% trehalose ($p < 0.01$). Trehalose substantially attenuated Aβ-induced inflammation of the hippocampus measured by the expression of microglial marker IBA1 ($p < 0.05$), and restored the density of Nissl-stained neurons in the frontal cortex and CA1 hippocampal area. All trehalose treatment regimens resulted in cognitive improvement of the fear-associated long-term memory and learning in the passive avoidance in the with the maximal effect of complete recovery for 4 % trehalose solution at continuous mode.

The disadvantage of the pharmacological model of AD is the short-term effect of neurointoxication, less pronounced signs of pathology, and spontaneous recovery due to the compensatory processes. At the same time, in humans, AD is a progressive disease which obviously needs prolonged treatment at different periods of disease course. We studied mice of the 5xFAD strain, which carry five mutations in two AD-related genes *App* and *Psen1* (*App* gene: K670N/M671L; I716V; V717I; P *Psen1* gene: M146L; L286V). At the age of 2 months, 5xFAD mice are characterized by the Aβ accumulation in the brain; at the age of 6 months they demonstrate phenotypic cognitive disorders [17]. In the T-maze and open field tests, no significant differences in the treatment of the three age-dependent groups of 5xFAD mice were observed. It is worth noting the normalizing effect on anxiety during prolonged (5 months) trehalose treatment. The core results of the therapy efficacy were obtained in the passive avoidance test. The main index (step-through latency) was sharply reduced in 5xFAD mice of all ages studied compared to the wild-type. Trehalose treatment was effective for all groups of mice of different ages ($p < 0.01$), the step-through latency was almost restored to the level of wild-type control mice. We can highlight the high results of behavioral recovery in the 2-month treatment group at the later period of the disease (5 months). Apparently, this is due to the beginning of the formation of phenotypic differences in the behavior of mice of this strain at this age [17].

Conclusion: 1. Elucidation of the dose-dependence of the therapeutic effect of trehalose showed that conventional treatment of Aβ-induced AD-like pathology in mice by drinking of 2 % trehalose solution could be improved by increasing the trehalose concentration (up to 4 %) in a regimen of constant (but not intermittent) trehalose consumption. In the subsequent experiments, the effective concentration of trehalose was increased to 3 %.

2. In an age-dependent study, the feasibility of treating a transgenic “progressive” model of AD with trehalose in young “latent” 2-month-old mice (corresponding to the onset of accumulation of the disease marker Aβ) was demonstrated. Similarly, treatment was also successful in mice of the mature 5-month-old age known for the appearance of phenotypic behavioral abnormalities. This latter result may be useful for considering the translational transition.

Funding: The study was supported by the Russian Science Foundation (grant No. 23-25-00393).

References

1. Basha F.H., Waseem M., Srinivasan H. *Cell Biochem Funct*. 2021;39(5):613-622
2. Lu Y., Li Z., Zhang S., Zhang T., Liu Y., Zhang L. *Theranostics*. 2023;13(2):736-766
3. Spilman P., Podlutskaya N., Hart M.J., Debnath J., Gorostiza O., Bredesen D., Richardson A., Strong R., Galvan V. *PLoS One*. 2010;5(4):e9979
4. Majumder S., Richardson A., Strong R., Oddo S. *PLoS One*. 2011;6(9):e25416
5. Sarkar S. Biochem Soc *Trans*. 2013;41(5):1103-1130
6. Switon K., Kotulska K., Janusz-Kaminska A., Zmorzynska J., Jaworski J. *Neuroscience*. 2017;341:112-153
7. Khalifeh M., Barreto G.E., Sahebkar A. *Neural Regen Res*. 2021;16(10):2026-2027
8. Pupyshev A.B., Klyushnik T.P., Akopyan A.A., Singh S.K., Tikhonova M.A. *Pharmacol Res*. 2022.183:106373
9. Yap K.H., Azmin S., Makpol S., Damanhuri H.A., Mustapha M., Hamzah J.C., Ibrahim N.M. *Neural Regen Res*. 2023;18(6):1179-1185
10. Perucho J., Casarejos M.J., Gomez A., Solano R.M., de Yebenes J.G., Mena M.A. *Curr Alzheimer Res*. 2012;9:334-343
11. Pupyshev A.B., Belichenko V.M., Tenditnik M.V., Bashirzade A.A., Dubrovina N.I., Ovsyukova M.V., Akopyan A.A., Fedoseeva L.A., Korolenko T.A., Amstislavskaya T.G., Tikhonova M.A. *Pharmacol Biochem Behav*. 2022;217:173406
12. Schaeffer V., Lavenir I., Ozcelik S., Tolnay M., Winkler D.T., Goedert M. *Brain*. 2012;135(7):2169-2177
13. He Q., Koprich J.B., Wang Y., Yu W.B., Xiao B.G., Brotchie J.M. Wang. *J. Mol. Neurobiol.* 2016;53(4):2258–2268.
14. Lin A.L., Zheng W., Halloran J.J., Burbank R.R., Hussong S.A., Hart M.J., Javors M., Shih Y.Y., Muir E., Solano Fonseca R., Strong R., Richardson A.G., Lechleiter J.D., Fox P.T., Galvan V. *J. Cereb. Blood Flow Metab*. 2013;33(9):1412-1421
15. Tikhonova M.A., Shoeva O.Y., Tenditnik,M.V., Ovsyukova M.V., Akopyan A.A., Dubrovina N.I., Amstislavskaya T.G., Khlestkina E.K. *Nutrients*. 2020;12(12)
16. Pupyshev A.B., Tikhonova M.A., Akopyan A.A., Tenditnik M.V., Dubrovina N.I., Korolenko T.A. *Pharmacol Biochem Behav*. 2019;177:1-11
17. Gorina Y.V., Vlasova O L., Bolshakova A.V., Salmina A.B. *I.M. Sechenov Russ. Physiol. J*. 2023;109(1):18-33 [in Russian]

doi 10.18699/bgrs2024-7.2-11

The effects of lymphocytes modulated *in vitro* by an original anticonvulsant on the brain functions in experimental alcoholism

Savkin I.V.*, Markova E.V.

Research Institute of Fundamental and Clinical Immunology, Laboratory of Neuroimmunology, Novosibirsk, Russia

**i.v.savkin2020@yandex.ru*

Key words: alcoholism, ortho-fluorobenzonal, lymphocytes, brain, behavior, neurotrophic factor, cytokines

Motivation and Aim: A growing body of literature indicates that the immune system plays a critical role in the development and maintenance of alcohol use disorder. Long-term alcohol consumption reduces the effectiveness of the cell-mediated and humoral immune response to infection and vaccination, which may be associated with the conversion of the naive T-lymphocyte phenotype, with a reduction in the number of CD4 and CD8 subpopulations of lymphocytes and modulation of their functional activity [1, 2]. GABAA-receptors (GABAA-R) represent the main inhibitory neurotransmitter system in the brain and play a central role in mediating the effects of ethanol. Changes in GABAA-R activity, similar to the effects on neuronal cells, cause modulation of the immune cells functional activity [2]. We first demonstrated that original compound ortho-fluoro-benzonal, artificial GABAA-R ligand, has immunostimulating properties and is able to restored long-term alcoholized mice lymphocytes activity *in vitro* through GABAA receptors [3-5]. Based on the previous results the purpose of the present study was to evaluate the effects of peripherally injected ortho-fluoro-benzonal modulated lymphocytes in in long-term alcoholized recipient.

Methods and Algorithms: Male (CBAxC57Bl/6)F1 mice in the state of alcohol dependence owing to 6-month 10 % ethanol exposure were administered intravenously with syngeneic splenic lymphocytes pre-cultivated with *ortho*-fluorobenzonal (10 μg/ml). The compound was synthesized in the research laboratory of drug synthesis of the National Research Tomsk Polytechnic University in the process of searching for highly effective anticonvulsants. The formation of the alcohol dependence was evaluated by a single injection of naloxone (3 mg/kg, subcutaneously) followed by visual assessment of the signs of "withdrawal syndrome". Recipient's consumption of 10 % ethanol solution under conditions of free choice with water was recorded daily at 10:00 for 7 days, starting from the first day after lymphocytes transplantation. Exploratory behavior was assessed in the "Open field" test. The content of cytokines and BDNF in brain structures of syngeneic long-term alcoholized recipients was assessed by ELISA using specific test systems: BenderMed Systems (Austria) for the determination of IFN-γ and IL-6; R&D Systems Inc. (United States) for the determination of IL-1β, IL-10, TNF-α and BDNF in accordance with the manufacturer's instructions. Cells staining was perfomed with CFSE.

Results: Daily consumption of ethanol in mice with alcohol dependence decreased starting from 2 days after intravenous administration of syngeneic lymphocytes precultivated with ortho-fluorobenzonal, which proved the alcohol motivation decrease.

Chronic exposure to ethanol resulted also in significant decrease in animals' behavioral activity in the "Open Field" test, while the intravenous administration of syngeneic lymphocytes precultivated with ortho-fluorobenzonal restored the motor and exploratory activities almost up to the intact healthy animal's level. Depressive-like behavior in chronic alcoholism is mediated by neuroinflammation. Alcohol is thought to alter immune signaling and increase neuroinflammation via two primary mechanisms: indirectly by initiating systemic production of proinflammatory cytokines; and directly through actions in the brain, whereby alcohol and potentially alcohol-induced neural damage stimulate the release of inflammatory molecules In particular, interaction of the key regulatory cytokine IL-1β with GABAA-R might account for the locomotory-depressive effects of this cytokine [1]. In the present study we revealed that above mentioned stimulation of behavioral activity in the "Open Field" test was recorded against the background of decreased levels of pro-inflammatory cytokines IL-1β, IL-6, TNF-α, IFN-γ in pathogenetically significant brain structures (frontal cortex, hippocampus and hypothalamus), which indicates a decrease in neuroinflammation. It is known that in chronic alcoholism neuroinflammation is associated with a decrease in the BDNF level, which causes in addition to neuroplasticity disorders the transition from occasional to compulsive alcohol consumption. We demonstrated a decrease BDNF level in the hippocampus of long-term alcoholized mice. After transplantation of lymphocytes modulated in vitro by the ortho-fluorobenzonal to syngeneic long-term alcoholized recipients an increase in the quantitative content of the hippocampal BDNF level was recorded, which indicates stimulation of neuroplasticity processes. There is also evidence that not only cytokines, but also immune cells can enter the brain and change the brain functions, including behavior, by direct contact with brain cells. The injected CFSE-labeled syngeneic lymphocytes modulated *in vitro* by ortho-fluorobenzonal 3 days after the systemic administration were recorded in the brain parenchyma of long-term alcoholized recipients. These lymphocytes were functionally active in the brain, which suggests, in particular, cell's direct influence on the brain functions.

Conclusion: Results demonstrated that transplantation of ortho-fluoro-benzonal-modulated lymphocytes caused positive psychoneuromodulating effect in long-term alcoholized recipients, manifested in editing behavioral patterns characteristic of alcoholism, reducing neuroinflammation and stimulating neuroplasticity.

Acknowledgements: This work was supported by the Russian federal budget allocated for the basic scientific research at the Federal State Budgetary Scientific Institution "Research Institute of Fundamental and Clinical Immunology". The authors are grateful to the T.V. Shushpanova (Tomsk National Research Medical Center) for providing the ortho-fluorobenzonal substance for the research.

References

1. Mereditha L.R, Burnettea E.M. et al. Immune treatments for alcohol use disorder: A translational framework. *Brain Behav Immun.* 2021;97:349-364. doi 10.1016/j.bbi.2021.07.023
2. Bjurstöm H. et al. GABA, a natural immunomodulator of T lymphocytes. *J Neuroimmunol*. 2008;205. (1-2):44-50. doi 10.1016/j.jneuroim.2008.08.017
3. Markova E.V. Immune cells and regulation of behavioral reactions in health and disease. Krasnoyarsk, 2021. (in Russian) doi 10.12731/978-5-907208-67-4
4. Markova E.V., Savkin I.V., Knyazheva M.A., Goldina I.A., Shushpanova T.V., Novozheeva T.P. Immunomodulator. Patent RU 2702114 C1, 2019 (in Russian)
5. Markova E., Savkin I., Shushpanova T. Effect of original anticonvulsant ortho-fluoro-benzonal on immune cells functional properties at chronic ethanol intoxication. *European Psychiatry*. 2022;65(S1): S816-S816. doi 10.1192/j.eurpsy.2022.2111

7 Симпозиум «Генетика/геномика, биоинформатика и системная биология животных»

Symposium “Genetics/genomics, bioinformatics and systems biology of animals”

doi 10.18699/bgrs2024-7.3-01

Полногеномный анализ ассоциаций с вариациями в уровне невербального интеллекта

Еникеева Р.[1, 2], Казанцева А.[1, 2, 3]*, Давыдова Ю.[1], Тахирова З.[2], Канапин А.[4], Мустафин Р.[5], Яковлева Д.[1, 2], Гилязова И.[1, 5], Малых С.[6, 7], Исматуллина В.[6], Хуснутдинова Э.[1, 2, 5]

[1] *Институт биохимии и генетики – обособленное структурное подразделение Уфимского федерального исследовательского центра РАН, Уфа, Россия*

[2] *Уфимский университет науки и технологий, Уфа, Россия*

[3] *Уфимский государственный нефтяной технический университет, Уфа, Россия*

[4] *Институт трансляционной биомедицины, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*

[5] *Башкирский государственный медицинский университет, Уфа, Россия*

[6] *Психологический институт Российской академии образования, Москва, Россия*

[7] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

* *enikeevarf@gmail.com*

Ключевые слова: GWAS; генетический вариант; невербальный интеллект

Мотивация и цель: В настоящее время выделяют множество типов интеллекта, но одним из самых устойчивых с точки зрения наследственности является невербальный интеллект (НВИ), который базируется на умении человека воспринимать и передавать информацию без слов, используя только визуальные образы. Исследования близнецов показали, что наследуемость невербального интеллекта составляет 64 % [1]. Однако генетическая структура невербального интеллекта сложна, гетерогенна и зависит от множества факторов. Это существенно усложняет изучение данного фенотипа, и современные генетические исследования невербального интеллекта не могут полностью объяснить его природу. Одним из самых эффективных методов изучения сложных когнитивных функций, в том числе и невербального интеллекта, является полногеномный анализ ассоциаций (GWAS), суть которого заключается в поиске однонуклеотидных полиморфизмов (SNP), которые чаще встречаются у людей с интересующим исследователя заболеванием или признаком. Целью данной работы стало проведение GWAS исследования вариаций в уровне невербального интеллекта в популяциях из Волго-Уральского региона.

Методы и алгоритмы: Выборка испытуемых состояла из студентов вузов Российской Федерации с известным уровнем НВИ, оцененным при помощи интернет-версии теста «Прогрессивные матрицы Равена» (N = 1422, 80 % – женщины; возраст 18–25 лет). Все участники эксперимента заявили, что у них нет психических заболеваний и они не наблюдаются у психиатра. Генотипирование было выполнено с использованием набора Infinium Global Screening Array v3.0 (Illumina, США). Статистический анализ, а также оценку качества генотипирования проводили с помощью программ Plink версии 1.09 (https://www.cog-genomics.org/plink/) и R (http://www.r-project.org6), данный анализ включал в себя коррекцию на этническую гетерогенность выборки с помощью включения главных компонент в качестве ковариат. Анализ ассоциации

проводился с применением разнообразных математических моделей, таких как общая линейная модель (GLM), смешанная линейная модель (MLM), а также модель с фиксированными и случайными эффектами (FarmCPU). Проверка на нормальность распределения показателей НВИ проводилась с использованием критерия Шапиро–Уилка. Для подтверждения достоверного влияния SNP и определения значимых регионов в геноме был применен тест Бонферрони (BFR) с пороговым значением $P \leq 0.05$ ($p < 5\times10^{-6}$). Исследование было одобрено этическим комитетом Института биохимии и генетики УФИЦ РАН.

Результаты: После проведения всех этапов контроля качества установлена ассоциация вариаций в уровне НВИ со следующими SNP: *rs35647108* ($p = 2.1\times10^{-6}$), *rs7745906* (p = 4.8×10^{-6}), *rs6908994* ($p = 5.0\times10^{-6}$), *rs9295957* ($p = 6.8\times10^{-6}$), *rs28360058* ($p = 7.3\times10^{-6}$), локализованные в близости от генов *HLA-B*, *HLA-C*, *rs7805604* ($p = 6.2\times10^{-6}$), *rs757911* ($p = 3.1\times10^{-5}$) в гене *GLI3*, *rs9323788* ($p = 3.5\times10^{-5}$) и *rs28381984* (в гене *MFF*, $p = 3.5\times10^{-5}$). Гены главного комплекса гистосовместимости HLA (Human Leukocyte Antigen), известного также как MHC (Major Histocompatibility Complex), располагаются на хромосоме 6, а сам комплекс играет ключевую роль в поддержании здоровья организма, устраняя чужеродные антигены. Интересно, что область HLA является наиболее полиморфной областью человеческого генома. Благодаря этому разнообразию система HLA обеспечивает максимальную защиту организма от широкого спектра патогенов, которые в противном случае могут оказывать негативное влияние на целостность нервной системы. Предыдущие исследования показали, что гены комплекса HLA могут оказывать влияние на возрастные изменения в структуре мозга, его функциях и когнитивных способностях у людей без деменции, в том числе при наличии рискового аллеля аполипопротеина *E4* (ген *APOE*), который ассоциирован со снижением когнитивных функций и развитием деменции. Это указывает на то, что комплекс HLA является значимым биологическим фактором, определяющим здоровье мозга [2]. Ген *GLI3*, расположенный на хромосоме 7 в регионе 7p14.1, играет важную роль в эмбриональном развитии. Он кодирует транскрипционный фактор, который участвует в регуляции экспрессии генов. Согласно данным литературных исследований, ген *GLI3* принимает участие в регуляции нейрогенеза на ранних этапах развития организма. Например, он участвует в развитии мозолистого тела – структуры переднего мозга, которая отвечает за передачу сенсорной, моторной и когнитивной информации между полушариями мозга. Кроме того, известно, что мутации в данном гене могут оказывать влияние на развитие мозга, вызывая аномалии в его структуре и функциях [3]. Третьим идентифицированным в рамках данного исследования геном является *MFF*, который кодирует белок, играющий ключевую роль в делении митохондрий, способствуя формированию новых митохондрий. Нарушение работы этого гена может привести к различным заболеваниям, связанным с энергетическим обменом клеток. Исследования показывают, что снижение экспрессии гена *MFF* может приводить к нарушению передачи сигналов в синапсах и препятствовать росту аксонов в процессе развития головного мозга [4].

Выводы: Таким образом, можно сделать вывод, что различия в уровне НВИ формируются при участии множества биологических систем, включая иммунную систему, систему энергетического обмена клетки и нейрогенеза. Тем не менее требуются дальнейшие исследования, целью которых будет подтверждение связи указанных SNP с уровнем НВИ в этнически схожей группе.

Genome-wide association study of variations in the level of nonverbal intelligence

Enikeeva R.[1, 2], Kazantseva A.[1, 2, 3]*, Davydova Y.[1], Takhirova Z.[2], Kanapin A.[4], Mustafin R.[5], Yakovleva D.[1, 2], Gilyazova I.[1, 5], Malykh S.[6, 7], Ismatullina V.[6], Khusnutdinova E.[1, 2, 5]

[1] *Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of RAS, Ufa, Russia*

[2] *Ufa University of Science and Technology, Ufa, Russia*

[3] *Ufa State Petroleum Technological University, Department of Molecular Technologies, Ufa, Russia*

[4] *Institute of Translational Biomedicine, Saint Petersburg State University, St. Petersburg, Russia*

[5] *Bashkir State Medical University, Ufa, Russia*

[6] *Psychological Institute of the Russian Academy of Education, Moscow, Russia*

[7] *Lomonosov Moscow State University, Moscow, Russia*

* *enikeevarf@gmail.com*

Key words: GWAS; genetic variant; non-verbal intelligence

Motivation and Aim: Currently, there are many types of intelligence, but one of the most stable from the point of view of heredity is non-verbal intelligence (NVI), which is based on a person's ability to perceive and transmit information without words, using only visual images. Twin studies have shown that the heritability of nonverbal intelligence is 64 % [1]. However, the genetic structure of nonverbal intelligence is complex, heterogeneous and depends on many factors. This significantly complicates the study of this phenotype, and modern genetic studies of nonverbal intelligence cannot fully explain its nature. One of the most effective methods for studying complex cognitive functions, including nonverbal intelligence, is genome-wide association study (GWAS), the essence of which is to search for single nucleotide polymorphisms (SNP), which are more common in people with a disease or trait of interest to the researcher. The purpose of this study was to conduct a GWAS study of variations in the level of non-verbal intelligence in populations from the Volga-Ural region.

Methods and Algorithms: The sample of subjects consisted of university students of the Russian Federation with a known level of NVI, assessed using the Internet version of the Raven Progressive Matrices test (N = 1422, 80 % were women; age 18–25 years). All participants in the experiment stated that they do not have mental illnesses, and they are not being monitored by a psychiatrist. Genotyping was performed using the Infinium Global Screening Array v3.0 kit (Illumina, USA). Statistical analysis, as well as an assessment of the quality of genotyping, was carried out using Plink software version 1.09 (https://www.cog-genomics.org/plink/) and R (http://www.r-project.org6), this analysis included correction for the ethnic heterogeneity of the sample by including the main components as covariates. The association analysis was carried out using a variety of mathematical models, such as the general linear model (GLM), the mixed linear model (MLM), as well as the fixed and random effects model (FarmCPU). Verification of the normality of the distribution of NVI indicators was carried out using the Shapiro–Wilk criterion. To confirm the reliable effect of SNP and identify significant regions in the genome, the Bonferroni test (BFR) with a threshold value of $P \leq 0.05$

($p < 5\times10^{-6}$) was used. The study was approved by the Ethics Committee of the Institute of Biochemistry and Genetics of the UFRC RAS.

Results: After carrying out all stages of quality control, an association of variations in the NVI level with the SNPs was established: *rs35647108* ($p = 2.1\times10^{-6}$), *rs7745906* ($p = 4.8\times10^{-6}$), *rs6908994* ($p = 5.0\times10^{-6}$), *rs9295957* ($p = 6.8\times10^{-6}$), *rs28360058* ($p = 7.3\times10^{-6}$), localized in proximity to the genes *HLA-B*, *HLA-C*, *rs7805604* ($p = 6.2\times10^{-6}$), *rs757911* ($p = 3.1\times10^{-5}$) in the *GLI3* gene, *rs9323788* ($p = 3.5\times10^{-5}$) and *rs28381984* (in the MFF gene, $p = 3.5\times10^{-5}$). The genes of the main histocompatibility complex HLA (Human Leukocyte Antigen), also known as MHC (Major Histocompatibility Complex) are located on the 6th chromosome, and the complex itself plays a key role in maintaining the health of the body by eliminating foreign antigens. Interestingly, the HLA region is the most polymorphic region of the human genome, due to this diversity, the HLA system provides maximum protection to the body from a wide range of pathogens that may otherwise have a negative impact on the integrity of the nervous system. Previous studies have shown that HLA complex genes can influence age-related changes in the structure of the brain, its functions, and cognitive abilities in people without dementia, including in the presence of the risky apolipoprotein *E4* allele (*APOE*), which is associated with cognitive decline and the development of dementia. This indicates that the HLA complex is a significant biological factor determining brain health [2]. The *GLI3* gene, located on chromosome 7 in the 7p14.1 region, plays an important role in embryonic development. It encodes a transcription factor that is involved in the regulation of gene expression. According to the data of literary studies, the *GLI3* gene participates in the regulation of neurogenesis at the early stages of the body's development. For example, it is involved in the development of the corpus callosum, a forebrain structure that is responsible for the transmission of sensory, motor, and cognitive information between the hemispheres of the brain. In addition, it is known that mutations in this gene can affect the development of the brain, causing abnormalities in its structure and functions [3]. The third gene identified in this study is *MFF*, which encodes a protein that plays a key role in mitochondrial division, contributing to the formation of new mitochondria. Disruption of this gene can lead to various diseases related to the energy metabolism of cells. Studies show that a decrease in the expression of the *MFF* gene can lead to impaired signal transmission in synapses and interfere with the growth of axons during brain development [4].

Conclusion: Thus, it can be concluded that differences in the level of HVI are formed with the participation of many biological systems, including the immune system, the system of energy metabolism of the cell and neurogenesis. However, further research is required to confirm the association of these SNPs with the level of HVI in an ethnically similar group.

Список литературы/References

1. Mustafin R.N., Kazantseva A.V., Enikeeva R.F., Malykh S.B., Khusnutdinova E.K. Longitudinal genetic studies of cognitive characteristics. *Vavilovskii Zhurnal Genet Selektsii.* 2020;24(1):87-95. doi 10.18699/VJ20.599
2. James L.M., Leuthold A.C., Georgopoulos A.P. Human leukocyte antigen (HLA) modulates the dependence on age of the variability of synchronous neural interactions. *Neurosci Insights.* 2023;18:26331055231159658. doi 10.1177/26331055231159658
3. Hussain I., Raza R.Z., Ali S., Abrar M., Abbasi A.A. Molecular signatures of selection on the human GLI3 associated central nervous system specific enhancers. *Dev Genes Evol.* 2021;231(1-2):21-32. doi 10.1007/s00427-021-00672-1
4. Nasca A., Nardecchia F., Commone A., Semeraro M., Legati A., Garavaglia B., Ghezzi D., Leuzzi V. Clinical and biochemical features in a patient with mitochondrial fission factor gene alteration. *Front Genet.* 2018;9:625. doi 10.3389/fgene.2018.00625

doi 10.18699/bgrs2024-7.3-02

Полногеномный анализ ассоциаций с уровнем агрессивности у лиц из Волго-Уральского региона

Казанцева А.[1, 2, 3]*, Давыдова Ю.[1], Еникеева Р.[1, 2], Тахирова З.[2], Канапин А.[4], Мустафин Р.[5], Яковлева Д.[1, 2], Гилязова И.[1, 5], Малых С.[6, 7], Исматуллина В.[6], Лобаскова М.[8], Хуснутдинова Э.[1, 2, 5]

[1] *Институт биохимии и генетики – обособленное структурное подразделение Уфимского федерального исследовательского центра РАН, Уфа, Россия*

[2] *Уфимский университет науки и технологий, Уфа, Россия*

[3] *Уфимский государственный нефтяной технический университет, Уфа, Россия*

[4] *Институт трансляционной биомедицины, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*

[5] *Башкирский государственный медицинский университет, Уфа, Россия*

[6] *Психологический институт Российской академии образования, Москва, Россия*

[7] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

[8] *Российская академия образования, Москва, Россия*

* *kazantsa@mail.ru*

Ключевые слова: GWAS; генетический вариант; агрессия; 14q31.3; 8p23.1

Мотивация и цель: Несмотря на важную роль агрессивности в выживании и адаптации человека с эволюционной точки зрения, агрессивное поведение (АП) представляет собой деструктивную форму коммуникации в случае снижения психоэмоционального контроля над ним [1]. Рассматривая манифестацию агрессивного и антисоциального поведения с позиций взаимодействия биологического (в том числе генетического) и социального компонентов [2], к настоящему времени проведены исследования, направленные на изучение определенных генетических вариантов в контексте их возможной связи с уровнем агрессивности как с использованием генно-кандидатного подхода [3], так и с помощью полногеномного анализа ассоциаций (GWAS) [4]. Несмотря на идентифицированные генетические локусы в данных работах, в своем большинстве исследуемым фенотипом являлось именно антисоциальное поведение, представляющее собой целый спектр различных синдромов и психопатологий: синдром дефицита внимания с гиперактивностью, расстройства аутистического спектра, зависимость от психоактивных веществ [4]. Однако GWAS работ, изучающих непосредственно уровень агрессивности у взрослых, к настоящему времени не опубликовано. Кроме того, особую актуальность представляет выявление генетических локусов с использованием GWAS анализа в российских когортах, поскольку все большее число работ свидетельствует об этноспецифичном характере ассоциаций, выявляемых в GWAS. В этой связи целью нашей работы являлось проведение полногеномного анализа ассоциаций с уровнем агрессивности в достаточной по объему выборке лиц из Волго-Уральского региона РФ.

Методы и алгоритмы: В качестве материала для исследования были использованы образцы ДНК (N = 1422) студентов вузов Республики Башкортостан и Удмуртской Республики (русские, татары, башкиры, удмурты, лица смешанной

этнической принадлежности; возраст 18–25 лет), не имевших психических нарушений. Генотипирование ~650 тыс. однонуклеотидных генетических локусов (SNP) в изученной выборке проводилось с использованием технологии ДНК-микрочипов и реагентов для полногеномного генотипирования образцов ДНК (Infinium Global Screening Array v3.0 Kit, Illumina). Уровень агрессивности был оценен с помощью опросника агрессии Басс–Перри (Buss–Perry Aggression Questionnaire, BPAQ-29). Полногеномный анализ ассоциации исследованных полиморфных локусов с уровнем агрессивности был осуществлен с коррекцией на этническую гетерогенность выборки с помощью включения главных компонент в качестве ковариат с использованием программного обеспечения Plink v.1.09 (https://www.cog-genomics.org/plink/), статистической среды R (http://www.r-project.org) и пакета прикладных программ GAPIT v.3 (Genome Association and Prediction Integrated Tool). Этот пакет программ позволяет проводить анализ ассоциаций с использованием различных математических моделей, включая общую линейную модель (GLM), смешанную линейную модель (MLM), модель с фиксированными и случайными эффектами (FarmCPU). Наряду с контролем качества массива данных на фенотипическом и генетическом уровнях была проведена математическая трансформация показателей для достижения соответствия Гауссовому распределению. Уровень статистической значимости принимался равным $p<5\times10^{-6}$ согласно предыдущим аналогичным исследованиям [5].

Результаты: В результате GWAS анализа были выбраны генетические локусы, пересекающиеся во всех трех проанализированных моделях (GLM, MLM, FarmCPU). Идентифицировано два SNP, ассоциированных с индивидуальными показателями в уровне агрессивности, включая rs17124581 ($p = 1.56\times10^{-6}$) и rs114798927 ($p = 3.97\times10^{-6}$), подходящих под выбранный уровень значимости. Генетический локус rs17124581 локализован в межгенной области, граничащей с генами *KCNK10* и *SPATA7*, а локус rs114798927 – рядом с генами *TNKS* и *PPP1R3B*. Локус rs17124581 находится в хромосомном регионе 14q31.3, значимость которого была показана для нормального нейронального развития, а изменения нуклеотидной последовательности в данном регионе были обнаружены ранее у пациентов с когнитивными нарушениями [6]. Ранее обнаруженный нами локус rs17124581 не был связан с развитием агрессивного или антисоциального поведения, хотя по результатам одного из GWAS данный локус также ассоциирован со скоростью обработки информации [7]. Учитывая, что одним из важных сдерживающих факторов развития агрессии является зрелость когнитивного функционирования, особенно таких отделов мозга, как префронтальная кора, можно предположить, что данный локус связан с развитием агрессивности посредством регуляции когнитивного функционирования. В непосредственной близости от локуса rs17124581 находятся гены, кодирующие белок семейства калиевых каналов (*KCNK10*) и *SPATA7*. Предыдущие GWAS работы подтверждают вовлеченность гена *KCNK10* в формирование никотиновой зависимости [8], которая зачастую коррелирует с манифестацией агрессивности. Кроме того, существуют данные, подтверждающие потенциальную связь гена *SPATA7* с развитием антисоциального поведения, включая шизофрению [9].

Другой локус, ассоциированный с агрессивностью в нашем исследовании, rs114798927, локализован в регионе 8p23.1, структурные изменения в котором в литературе связывают с манифестацией рискового поведения [10].

Выводы: Впервые в рамках GWAS, проведенного на российской когорте лиц 18–25 лет из Волго-Уральского региона, были выявлены генетические локусы rs17124581 и rs114798927, ассоциированные с уровнем агрессивности. Полученный результат свидетельствует в пользу возможной вовлеченности генов калиевых каналов (*KCNK10*) и танкиразы (*TNKS*) в формирование агрессивности как посредством участия в эпигенетической регуляции других молекулярных компонентов, так и с помощью регуляции когнитивного контроля. Несмотря на достаточный по размеру объем испытуемых и введение основных главных компонент для коррекции на этническую принадлежность в качестве ковариат в математические модели, в настоящей работе не были учтены различные средовые факторы, потенциально влияющие на манифестацию агрессивности. Дальнейшие исследования в этой области должны включать проведение репликативного анализа того же фенотипа с целью верификации идентифицированных генетических локусов как в выборке лиц из того же региона, так и из других регионов России.

Финансирование: Исследование выполнено в рамках государственного задания Министерства науки и высшего образования Российской Федерации (№ 122041400169-2). Образцы ДНК взяты из ЦКП «Коллекция биологических материалов человека» ИБГ УФИЦ РАН, поддержанного Программой биоресурсных коллекций ФАНО России (№ 007-030164/2).

Genome-wide association study of aggression level in individuals from the Volga-Ural region

Kazantseva A.[1, 2, 3]*, Davydova Y.[1], Enikeeva R.[1, 2], Takhirova Z.[2], Kanapin A.[4], Mustafin R.[5], Yakovleva D.[1, 2], Gilyazova I.[1, 5], Malykh S.[6, 7], Ismatullina V.[6], Lobaskova M.[8], Khusnutdinova E.[1, 2, 5]

[1] *Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of RAS, Ufa, Russia*

[2] *Ufa University of Science and Technology, Ufa, Russia*

[3] *Ufa State Petroleum Technological University, Department of Molecular Technologies; Ufa, Russia*

[4] *Institute of Translational Biomedicine, Saint Petersburg State University, St. Petersburg, Russia*

[5] *Bashkir State Medical University, Ufa, Russia*

[6] *Psychological Institute of the Russian Academy of Education, Moscow, Russia*

[7] *Lomonosov Moscow State University, Moscow, Russia*

[8] *Russian Academy of Education, Moscow, Russia*

* *kazantsa@mail.ru*

Key words: GWAS; SNP; aggression; 14q31.3; 8p23.1

Motivation and Aim: Despite the important role of aggression in human survival and adaptation from an evolutionary perspective, aggressive behavior (AB) is a destructive form of communication developed under reduced emotional control [1]. Considering the manifestation of aggressive and antisocial behavior as a result of interaction of biological (including genetic) and social components [2], the studies conducted to date were directed to the search of certain genetic variants related to aggression using both a candidate gene approach [3] and genome-wide association studies (GWAS) [4]. Genetic

loci, which were identified in these studies, were precisely related to the development of antisocial behavior, which included a whole range of different syndromes and psychopathologies such as attention deficit hyperactivity disorder, autism spectrum disorders, and addiction [4]. However, to date no GWAS related to the aggression level itself rather than antisocial behavior have been published. In addition, identification of SNPs via GWAS in Russian cohorts is of particular relevance since an increasing number of studies indicate an ethnicity specific nature of associations identified in GWAS. In this regard, the present study aimed to perform GWAS of aggression level in a moderate sample of young adults from the Volga-Ural region of the Russian Federation.

Methods and Algorithms: DNA samples (N = 1422) obtained from University students of the Republic of Bashkortostan and the Udmurt Republic (Russians, Tatars, Bashkirs, Udmurts, persons of mixed ethnicity; aged 18–25 years), who were absent of any mental disorder, were used as a material for the study. Genotyping of ~ 650k SNPs in the studied sample was carried out using DNA microarray technology and reagents (Infinium Global Screening Array v3.0 Kit, Illumina). Aggression level was assessed using the Buss-Perry Aggression Questionnaire (BPAQ-29). GWAS of aggression level was carried out with correction for ethnic heterogeneity of the sample by including the main components as covariates using Plink v.1.09 (https://www.cog-genomics.org/plink /), R (http://www.r-project.org) and GAPIT v.3 package (Genome Association and Prediction Integrated Tool). This software enables to analyze associations using various mathematical models, including the general linear model (GLM), the mixed linear model (MLM), the fixed and random effects model (FarmCPU). Along with the quality control of phenotype and genotypes, a mathematical transformation of aggression level was carried out to achieve compliance with the Gaussian distribution. The level of statistical significance was assumed at $p<5\times10^{-6}$ according to previous study [5].

Results: As a result of the GWAS analysis we have selected SNPs intersected in all three examined models (GLM, MLM and FarmCPU) as significant ones. Two SNPs associated with individual aggression level were detected, including rs17124581 ($p = 1.56\times10^{-6}$) and rs114798927 ($p = 3.97\times10^{-6}$). The rs17124581 is located in the intergenic region neighboring the *KCNK10* and *SPATA7* genes, and rs114798927 is located near the *TNKS* and *PPP1R3B* genes. The rs17124581 resides 14q31.3 region, which was shown to be important for normal neuronal development, and nucleotide substitutions in this region were previously reported in patients with cognitive impairment [6]. The rs17124581 has not been previously associated with aggressive or antisocial behavior; however, based on the data from previous GWAS, this SNP was also associated with information processing speed [7]. Considering that cognitive emotional control is one of the important restraining factors for developing aggression, including especially prefrontal cortex, it can be assumed that identified SNP is associated with aggression through the regulation of cognitive functioning. The genes encoding the potassium channel family protein (*KCNK10*) and *SPATA7* are located in a close proximity to rs17124581. Previous GWAS confirm the involvement of the *KCNK10* gene in nicotine dependence [8], which often correlates with the aggression level. In addition, there is evidence confirming the potential association of the *SPATA7* gene with antisocial behavior, including schizophrenia [9].

Another SNP associated with aggression in our study, rs114798927, is located in 8p23.1 region, structural changes in which were associated with risky behavior [10].

Conclusion: For the first time, we identified rs17124581 and rs114798927 as associated with aggression level in mentally healthy young adults from Russia aged 18–25 years,

which implicated GWAS framework. The obtained results indicate the possible involvement of potassium channel genes (*KCNK10*) and tankyrase (*TNKS*) in the development of human aggression via both epigenetic regulation of other molecular components and regulation of cognitive control. Despite the sufficient sample size and control for the principal components to correct for ethnicity in mathematical models, various environmental factors potentially affecting manifestation of aggression were non-considered in the present study. Further research in this field should represent a replication analysis in order to verify the association of identified SNPs with the same phenotype, which has to be carried out in individuals from the same region and from other regions of Russia.

Funding: The study is supported by the State Contract of the Ministry of Science and Higher Education of RF (No. 122041400169-2). The samples were obtained from the CCU "Collection of human biological materials" of IBG UFRC RAS supported by the Program of Bioresources collections of FASO of RF (No. 007-030164/2).

Список литературы/References

1. Baron R.A., Richardson D.R. Human aggression. New York: Springer, 2004
2. Kazantseva A.V. et al. Antisocial behavior: nature vs. nurture. *Opera Med Physiol*. 2023;10(3):83-94. doi 10.24412/2500-2295-2023-3-83-94
3. Davydova Y.D. et al. The role of oxytocin receptor (OXTR) gene polymorphisms in the development of aggressive behavior in healthy individuals. *Russ J Genet.* 2020;56(9):1129-1138. doi 10.1134/S1022795420090057
4. Tielbeek J.J. et al. Genome-Wide association studies of a broad spectrum of antisocial behavior. *JAMA Psychiatry*. 2017;74(12):1242-1250. doi 10.1001/jamapsychiatry.2017.3069
5. Levchenko A. et al. A genome-wide association study identifies a gene network associated with paranoid schizophrenia and antipsychotics-induced tardive dyskinesia. *Prog Neuropsychopharmacol. Biol Psychiatry*. 2021;105:110134
6. Chen C.P. et al. A paternally derived inverted duplication of distal 14q with a terminal 14q deletion. *Am J Med Genet A*. 2005;139A(2):146-150. doi 10.1002/ajmg.a.30997
7. Luciano M. et al. Whole genome association scan for genetic polymorphisms influencing information processing speed. *Biol Psychol.* 2011;86(3):193-202. doi 10.1016/j.biopsycho.2010.11.008
8. Zhou H. et al. Genome-wide association study identifies glutamate ionotropic receptor GRIA4 as a risk gene for comorbid nicotine dependence and major depression. *Transl Psychiatry*. 2018;8(1):208. doi 10.1038/s41398-018-0258-8
9. Igeta H. et al. Rare compound heterozygous missense SPATA7 variations and risk of schizophrenia; whole-exome sequencing in a consanguineous family with affected siblings, follow-up sequencing and a case-control study. *Neuropsychiatr Dis Treat*. 2019;15:2353-2363. doi 10.2147/NDT.S218773
10. Karlsson Linnér R. et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet*. 2019;51(2):245-257. doi 10.1038/s41588-018-0309-3

doi 10.18699/bgrs2024-7.3-03

Влияние функционального нокаута гена *Ptpn5*, кодирующего стриатумспецифичную протеинтирозин-фосфатазу STEP, на поведение мышей

Москалюк В.С.*, Комлева П.Д., Хоцкин Н.В., Базовкина Д.В., Куликова Е.А.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
** moskaliukvs@bionet.nsc.ru*

Ключевые слова: стриатумспецифичная протеинтирозинфосфатаза STEP; нокаут; нейропластичность; поведение; мыши

Мотивация и цель: Стриатумспецифичная протеинтирозинфосфатаза STEP, кодируемая геном *Ptpn5*, участвует в ключевых сигнальных каскадах нервной клетки, регулирующих процессы нейропластичности, и связана с глутамат-, дофамин- и серотонинергическими системами мозга. Известна роль данной фосфатазы в патогенезе ряда нейродегенеративных и психических заболеваний, таких как болезни Альцгеймера, Паркинсона, Хантингтона, а также шизофрении и депрессии [1]. Более того, ингибитор белка STEP оказывал анксиолитическое, антиагрессивное и антидепрессантное действия [2, 3], а также корректировал поведение животных в моделях нейродегенеративных заболеваний. Эти данные указывают на возможное участие фосфатазы STEP в регуляции когнитивных функций и поведения. Целью работы было исследовать влияние функционального нокаута гена *Ptpn5* на поведение мышей.
Методы и алгоритмы: В работе использовались половозрелые (возраст 2 месяца) самцы мышей линии C57BL/6 и мыши с нокаутом гена *Ptpn5* (STEP KO, *Ptpn5*$^{-/-}$). У линии мышей STEP KO делеция в гене *Ptpn5* приводит к разрушению фосфатазного домена в продукте гена. Поведение животных оценивалось в тестах: открытое поле, приподнятый крестообразный лабиринт, закапывание шариков, новый объект, ротарод, водный лабиринт Морриса, трехкамерный социальный тест и тест на социальное взаимодействие.
Результаты: В тесте «открытое поле» мыши с нокаутов гена *Ptpn5* не отличались от мышей дикого типа по двигательной и исследовательской активности, но в тесте «приподнятый крестообразный лабиринт» они проводили больше времени в открытых рукавах, что говорит о сниженной тревожности, и чаще свешивались с открытых рукавов, что указывает на повышенную исследовательскую активность. Мыши STEP KO закапывали меньше шариков, что говорит о сниженном стереотипном поведении. В тесте «водный лабиринт Морриса» на пространственное обучение группе мышей с нокаутом в фазе обучения требовалось больше времени, чтобы найти платформу, а на пятый контрольный день они проводили меньше времени в целевой зоне, чем мыши дикого типа. Это говорит о пагубном влиянии дисфункции фосфатазы STEP на пространственную память. Нами не было выявлено влияния нокаута гена *Ptpn5* на социальное поведение, на моторную активность в тесте «ротарод» и на память в тесте «новый объект».
Выводы: Таким образом, функциональный нокаут гена *Ptpn5* увеличил исследовательское и снизил тревожное и стереотипное поведение мышей. В то же

время дисфункция фосфатазы STEP пагубно отразилась на пространственном обучении и памяти.
Финансирование: Исследования выполнены при поддержке бюджетного проекта FWNR-2022-0010.

Effects of functional knockout of *Ptpn5* gene, encoding striatal-enriched protein tyrosine phosphatase, on behavior of mice

Moskaliuk V.S.*, Komleva P.D., Khotskin N.V., Bazovkina D.V., Kulikova E.A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** moskaliukvs@bionet.nsc.ru*

Key words: striatal-enriched protein tyrosine phosphatase STEP; knockout; neuroplasticity; behavior; mice

Motivation and Aim: Striatal-enriched protein tyrosine phosphatase STEP, encoded by the *Ptpn5* gene, is one of the key signalling molecules regulating neuroplasticity and is connected to the glutamate-, dopamine- and serotoninergic systems. STEP is known to take part in the pathogenesis of a number of neurodegenerative and mental diseases such as Alzheimer's, Parkinson's, Huntington's, schizophrenia and depression [1]. Furthermore, the STEP inhibitor exerted anxiolytic, antiaggressive and antidepressant actions [2, 3], as well as corrected pathological behavior in animal models of neurodegenerative diseases. These data indicate possible involvement of STEP phosphatase in the regulation of cognitive functions and behaviors. The goal of this study was to investigate the effect of functional knockout of the *Ptpn5* gene on the behavior of mice.
Methods and Algorithms: The study was conducted on mature (2-month-old) male mice of the C57BL/6 strain and mice with the knockout of the *Ptpn5* gene (STEP KO, *Ptpn5*$^{-/-}$). In the STEP KO mice, the deletion in the *Ptpn5* gene leads to the destruction of the phosphatase domain in the resulted protein. Animal behavior was assessed in following tests: open field, elevated plus maze, marble burial, novel object, rotarod, Morris water maze, three-chamber test and social interaction test.
Results: In the open field test mice with the knockout of the *Ptpn5* gene did not differ from wild type mice in motor and research activity, but in the elevated plus maze test they showed reduced anxiety-like behavior and increased research activity. STEP KO mice buried fewer marbles, suggesting decreased stereotypical behavior. In the Morris water maze, a group of mice with a knockout on average took longer to find the platform, and on the 5th test day they spent less time in the target area than the wild type mice. This suggests that STEP phosphatase dysfunction has a detrimental effect on spatial memory. We have not found any influence of the *Ptpn5* gene knockout on social behavior, motor activity in the rotarod test and memory in the novel object test.
Conclusion: Thus, functional knockout of the *Ptpn5* gene increased the research activity and reduced anxiety and stereotypical behavior in mice. At the same time, the dysfunction of STEP phosphatase adversely affected spatial learning and memory.
Funding: This work was supported by a Basic Research Project for a Young Researcher FWNR-2022-0010.

Список литературы/References

1. Mahaman Y.A.R., Huang F., Embaye K.S., Wang X., Zhu F. The implication of STEP in synaptic plasticity and cognitive impairments in alzheimer's disease and other neurological disorders. *Front Cell Dev Biol*. 2021;9:680118
2. Moskaliuk V.S., Kozhemyakina R.V., Bazovkina D.V. et al. On an association between fear-induced aggression and striatal-enriched protein tyrosine phosphatase (STEP) in the brain of Norway rats. *Biomed Pharmacother*. 2022;147:112667
3. Kulikova E.A., Khotskin N.V., Illarionova N.B. et al. Inhibitor of striatal-enriched protein tyrosine phosphatase, (TC-2153), produces antidepressant-like effect and decreases functional activity and protein Level of 5-HT2A receptor in the brain. *Neuroscience*. 2018;394:220-231

doi 10.18699/bgrs2024-7.3-04

Дофаминергическая система мозга крыс в механизмах генетически детерминированной защитно-оборонительной агрессии по отношению к человеку

Правикова П.*, Москалюк В., Базовкина Д., Науменко В.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
** PollyPravi@yandex.ru*

Ключевые слова: генетически детерминированная агрессия; дофаминергическая система мозга; рецепторы дофамина; дофамин; гомованилиновая кислота; гипоталамус; крысы

Мотивация и цель: В продолжение эксперимента Д.К. Беляева по доместикации лисиц в 1970-х гг. П.М. Бородиным было предпринято исследование по целенаправленному «одомашниванию» серых крыс (*Rattus norvegicus*), результатом которого стало выведение ручной линии и крыс с генетически детерминированной защитно-оборонительной агрессией. Полученные две линии крыс с различной степенью агрессивности по отношению к человеку являются удобной моделью для изучения одной из основных проблем современной нейронауки – механизмов формирования агрессивного поведения. Моноаминовые нейротрансмиттерные системы – перспективная мишень в исследовании механизмов регуляции и формирования агрессивности. Вместе с тем, несмотря на большой массив данных о вовлечении серотониновой системы мозга в регуляцию различных форм агрессии [1], на сегодняшний день крайне мало информации о роли дофаминергической системы в поддержании защитно-оборонительного поведения. Цель данной работы – исследование дофаминовой системы в структурах мозга крыс с генетически детерминированной агрессией по отношению человеку или ее отсутствием.
Методы и алгоритмы: Анализ проведен на взрослых самцах крыс (*Rattus norvegicus*), селекционированных в течение 98 поколений на высокую агрессию по отношению к человеку и ее отсутствие ($n = 8$ для каждой линии). Исследовался средний мозг, гипоталамус, гиппокамп, стриатум и фронтальная кора. Уровень дофамина (ДА), его метаболитов 3,4-дигидроксифенилуксусной кислоты (DOPAC) и гомованилиновой кислоты (ГВК) определялся в исследуемых структурах мозга методом ВЭЖХ. Индекс метаболизма ДА вычислялся как соотношение DOPAC или ГВК к ДА. Методом ОТ-ПЦР в реальном времени оценивался уровень мРНК основных дофаминовых рецепторов (*Drd1*, *Drd2*), транскрипционного фактора *Creb1*, ДА-транспортера (*Slc6a3*), тирозингидроксилазы (*Th*) и катехол-О-метилтрансферазы (*Comt*).
Результаты: Существенные межлинейные изменения в уровне экспрессии ДА-рецепторов были выявлены лишь в гиппокампе: у агрессивных крыс зафиксировано увеличение уровня мРНК *Drd2*, кодирующего ДА-рецептор 2-го типа (D2R). При этом выявленное увеличение экспрессии D2R, трансдукция сигнала по G-зависимому пути которого приводит к подавлению цАМФ-сигнального каскада [2], сопровождалось ростом уровня мРНК *Creb1*, кодирующего ассоциированный белок к цАМФ элементу, что, вероятно, является

компенсаторным эффектом поддержания стабильности ДА-нейротрансмиссии. Установлена прямая корреляционная связь между уровнем мРНК *Drd2* и *Creb1* ($r = 0.716$; $p < 0.002$). В гиппокампе, как и в стриатуме, не обнаружено межлинейных различий в уровне ДА и его метаболитов (DOPAC, ГВК). В среднем мозге у агрессивных крыс по сравнению с ручной линией на фоне увеличения уровня экспрессии гена *Th*, кодирующего основной фермент синтеза ДА, установлен рост уровня DOPAC, при этом концентрация ДА и ГВК не изменилась. Во фронтальной коре у агрессивной линии крыс выявлено увеличение концентрации ДА вследствие, по-видимому, межлинейных различий в уровне мРНК *Th*, которая у ручных крыс не детектировалась вовсе. Вместе с тем самые значимые межлинейные различия в уровне метаболизма ДА установлены нами в гипоталамусе, в структуре, как известно, ответственной за регуляцию агрессивного поведения [3]. У агрессивной линии крыс в гипоталамусе, несмотря на отсутствие изменений в уровне экспрессии генов метаболизма ДА (*Th*, *Comt*), зафиксировано снижение в уровне ДА на фоне существенного роста его конечного метаболита ГВК, при этом концентрация DOPAC, напротив, снижалась.

Выводы: Таким образом, нами были впервые установлено существенное увеличение уровня метаболизма ДА в гипоталамусе агрессивных крыс. Исходя из полученных результатов, а также с учетом ранее выдвинутой гипотезы об изменчивости гипоталамо-гипофизарно-надпочечниковой системы (ГГНС) в результате отбора на различную степень агрессивности по отношению к человеку [4] можно предположить участие гипоталамической дофаминергической системы в механизмах формирования защитно-оборонительного поведения.

Финансирование: Содержание животных осуществлялось за счет средств бюджетного проекта FWNR-2022-0023. Работа выполнена при поддержке Российского научного фонда (грант № 22-15-00028).

Dopaminergic system of the rats brain in the mechanisms of genetically defined defensive aggression towards man

Pravikova P.*, Moskaliuk V., Bazovkina D., Naumenko V.

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

** PollyPravi@yandex.ru*

Key words: genetically defined aggression; dopaminergic system; dopamine receptors; dopamine; homovanillic acid; hypothalamus; rats

Motivation and Aim: In continuation of D.K. Belyaev's experiment on fox domestication, in the 1970s P.M. Borodin undertook a research on the goal-oriented "domestication" of Norway rats (*Rattus norvegicus*), which resulted in the breeding of a tame strain and rats with genetically defined defensive aggression. The resulting two strains of rats with different degrees of aggressiveness towards humans are a useful model for investigating one of the main problems of modern neuroscience, namely the mechanisms of aggressive behavior formation. Monoamine neurotransmitter systems are a promising target in the analysis of the regulation and formation mechanisms of aggressiveness. At the same time, despite a large amount of data on the involvement of the brain serotonin system in the regulation of various aggression forms [1], today there is lack of information about the role of the dopaminergic system in maintaining defensive

behavior.The aim of this work is to investigate the dopamine system in the brain structures of rats with genetically defined aggression towards man or its absence.

Methods and Algorithms: The analysis was carried out on adult male rats (*Rattus norvegicus*), selected (for 98 generations) for high defensive aggression towards man and its absence (n = 8 for each strain). The midbrain, hypothalamus, hippocampus, striatum and frontal cortex were studied.The levels of dopamine (DA), its metabolites 3,4-dihydroxyphenylacetic acid (DOPAC) and homovanillic acid (HVA) were determined in the investidated brain structures by HPLC method. The index of brain DA-metabolism was calculated as the ratio of DOPAC or HVA to DA. Real-time RT-qPCR was used to evaluate the mRNA levels of the main DA receptors (*Drd1*, *Drd2*), transcription factor *Creb1*, DA transporter (*Slc6a3*), tyrosine hydroxylase (*Th*) and catechol-O-methyltransferase (*Comt*).

Results: Significant interstrain changes in the expression level of DA receptors were detected only in the hippocampus; in aggressive rats, an increase in the *Drd2* mRNA level, encoding dopamine D2 receptor (D2R), was recorded. At the same time, the detected increase in the *Drd2* gene expression, that G-dependent signal transduction leads to the suppression of the cAMP-cascade [2], was accompanied by an increase in the *Creb1* mRNA level, encoding an cAMP-response element binding protein, which is probably a compensatory effect of maintaining the stability of DA neurotransmission. A positive correlation was showed between the *Drd2* and *Creb1* mRNA levels (r = 0.716; $p < 0.002$). In the hippocampus, as in the striatum, no interstrain differences were found in the DA and its metabolites (DOPAC, HVA) levels. In the midbrain of aggressive rats, compared with the tame strain, with an increase in the expression level of the *Th* gene, encoding the main enzyme for the synthesis of DA, an increase in the level of DOPAC was established, while the concentration of DA and HVA did not change. In the midbrain of aggressive rats, compared with the tame strain, there was an increase in the *Th* gene expression level, encoding the main enzyme for the DA synthesis, whereas the concentration of DA and HVA in this structure did not change, while an increase in the level of DOPAC was found. In the frontal cortex of the aggressive strain, an increase in DA concentration was detected due, apparently, to interstrain differences in the *Th* mRNA level, which was not detected at all in the tame rats. With that, we revealed the most significant interline differences in the level of DA metabolism in the hypothalamus, a structure known to be responsible for regulation aggressive behavior [3]. In an aggressive strain of rats, in the hypothalamus, despite the absence of changes in the DA metabolism genes (*Th*, *Comt*) expressionl levels, a decrease in the DA level was recorded against the background a significant increase in the final DA metabolite (HVA), while the concentration of DOPAC, on the contrary, decreased.

Conclusion: Thus, we were the first to reveal a significant increase in the hypothalamic DA metabolism level of aggressive rats. Based on the results obtained, and also taking into account the previously proposed hypothesis about the variability of the hypothalamic-pituitary-adrenal (HPA) axis as a result of selection for different degrees of aggressiveness towards man [4], we can assume the participation of the hypothalamic dopaminergic system in the mechanisms of defensive behavior formation.

Funding: Animals maintenance was provided by the Budget Project FWNR-2022-0023. This work was supported by the Russian Science Foundation (project No. 22-15-00028).

Список литературы/References

1. Olivier B. Serotonin and aggression. *Ann NY Acad Sci.* 2004;1036:382-392. doi 10.1196/annals.1330.022
2. Bonci A., Hopf F.W. The dopamine D2 receptor: new surprises from an old friend. *Neuron.* 2005;47(3):335-338. doi 10.1016/j.neuron.2005.07.015
3. Bermond B., Mos J., Meelis W., van der Poel A.M., Kruk M.R. Aggression induced by stimulation of the hypothalamus: effects of androgens. *Pharmacol Biochem Behav.* 1982;16(1):41-45. doi 10.1016/0091-3057(82)90010-7
4. Naumenko E.V., Popova N.K., Nikulina E.M., Dygalo N.N., Shishkina G.T., Borodin P.M., Markel A.L. Behavior, adrenocortical activity, and brain monoamines in Norway rats selected for reduced aggressiveness towards man. *Pharmacol Biochem Behav.* 1989;33(1):85-91. doi 10.1016/0091-3057(89)90434-6

doi 10.18699/bgrs2024-7.3-05

Анализ механизмов большого депрессивного расстройства на основе интеллектуального анализа текстов

Тарасова О.[1]*, Бизюкова Н.[1], Сухачёв В.[1], Иванов С.[1, 2], Филимонов Д.[1], Поройков В.[1]

[1] *Институт биомедицинской химии им. В.Н. Ореховича, Москва, Россия*

[2] *Российский национальный исследовательский медицинский университет им. Н.И. Пирогова, Москва, Россия*

* *olga.a.tarasova@gmail.com*

Ключевые слова: большое депрессивное расстройство; патогенез; интеллектуальный анализ текстов; эпигенетическая регуляция; антидепрессанты

Мотивация и цель: Депрессивные расстройства приводят к снижению качества жизни миллионов людей. Совокупность факторов, ведущих к возникновению и развитию этого психического заболевания, включает в себя генетическую предрасположенность, факторы среды (в том числе стресс и травмы), а также наличие соматических заболеваний [1]. В недавних исследованиях было показано, что ранняя психическая травматизация ребенка приводит к изменению эпигенетической регуляции генов [2], что может оказывать влияние на развитие ряда психических заболеваний, включая большое депрессивное расстройство (БДР). Целью нашей работы является исследование совокупности причин большого депрессивного расстройства с применением методов интеллектуального анализа текстов.

Методы и алгоритмы: С применением ранее разработанных алгоритмов отбора аннотаций релевантных научных публикаций, извлечения наименований биологических и химических объектов из их текстов и анализа взаимосвязи между извлеченными объектами [3, 4] нами была выявлена совокупность наименований генов, белков, микроРНК, которые потенциально ассоциированы с БДД, и определен ряд молекулярных механизмов этого заболевания, а также причин, не относящихся к изменениям на молекулярном уровне. Отбор релевантных публикаций производился на основе ключевых слов по запросу в БД библиографической информации PubMed. Распознавание наименований проводилось с применением метода на основе условных случайных полей (CRF) [4], библиотеки HunFlair для языка Python [5], регулярных выражений и словарей. Исследование связей между терминами проводилось на основе классификации предложений, содержащих и не содержащих указание на наличие взаимосвязи между терминами, с использованием наивного байесовского классификатора, применение которого к задаче распознавания наименований было исследовано нами ранее [3].

Результаты: По результатам запросов в БД PubMed было извлечено свыше 3000 записей, соответствующих аннотациям текстов публикаций. По результатам извлечения ассоциаций между объектами наименований белков, генов и микроРНК и их влияния на развитие БДР были идентифицированы гены дофаминергической и серотонинергической систем, глюкокортикоидного рецептора, изменение уровня экспрессии которых может быть обусловлено

воздействием неблагоприятных обстоятельств (например, сенсорная депривация) в раннем детском возрасте и, таким образом, может непосредственно влиять на развитие депрессии при воздействии неблагоприятных факторов во взрослом возрасте (формирование порочного круга). Дополнительно, были выявлены взаимосвязи между эпигенетическими механизмами регуляции рецепторов эстрогена, которые могут вносить вклад в развитие психических заболеваний. При распознавании в текстах наименований факторов, которые вносят вклад в развитие БДР, но не могут быть отнесены к молекулярным механизмам развития депрессии, нами были найдены также такие факторы, как тип привязанности ребенка, наличие в его близком окружении благоприятного объекта (близкого взрослого), а также наличие сопутствующих расстройств, таких как пограничное расстройство личности или тревожно-депрессивные расстройства, которые могут как сопровождаться более тяжелой симптоматикой, так и обосновывать необходимость дифференциальной диагностики у пациентов с БДР.
Выводы: С применением интеллектуального анализа текстов исследованы потенциальные механизмы развития большого депрессивного расстройства. Выявленные причины включают в себя совокупность биологических и психологических факторов. Некоторые из найденных молекулярных механизмов, такие как регуляция генов дофаминергической системы, подтверждены в более ранних исследованиях, в то время как идентифицированные взаимосвязи между регуляцией эстрогенового и глюкокортикоидного рецепторов и развитием депрессии у взрослых, а также оценка соотношения психологических и наследственных факторов в динамике у пациентов с большим депрессивным расстройством представляют интерес.
Финансирование: Работа выполнена при поддержке гранта Российского научного фонда № 24-25-00453.

An analysis of major depressive disorders using automated text analysis

Tarasova O.[1]*, Biziukova N.[1], Sukhachev V.[1], Ivanov S.[1, 2], Filimonov D.[1], Poroikov V.[1]

[1] *Institute of Biomedical Chemistry, Moscow, Russia*

[2] *Pirogov Russian National Research Medical University, Moscow, Russia*

* *olga.a.tarasova@gmail.com*

Key words: major depressive disorders; text mining; pathogenesis; epigenetic regulation; antidepressants

Motivation and Aim: Major depressive disorder (MDD) can have a significant impact on the quality of life for millions of people. The development of MDD may be influenced by genetic and environmental factors, such as stress and trauma, as well as the co-morbidities of a patient [1]. In recent studies, an influence of early childhood psychological trauma on epigenetic gene regulation was shown [2]. The purpose of our study is to examine the impact of several biological and psychological factors on mental disorders, such as MDD, using text-mining approach.
Methods and Algorithms: Based on earlier developed algorithms of extracting biological and chemical named entities from the texts of scientific publications [3, 4] we carried

out named entity recognition of objects from their texts and analyzing the relationship between the extracted objects. We identified a set of genes, proteins and microRNAs that are associated with MDD and may be involved in the molecular mechanisms of the disorder. We selected relevant texts of publications based on keywords such as “major depressive disorder and epigenetics regulation”, “major depressive disorder and co-morbidities”, “major depressive disorder and genetic polymorphisms”. Named entity recognition was carried out based on the CFR [4] and HunFlair [5] packadges. The relationships between the terms were extracted based on the classification of sentences into those that contain and do not contain an indication of a relationship between terms, using a naive Bayesian classifier. We demonstrated the applicability of naïve Bayesian classifier for the task of name recognition earlier [3].

Results: Based on queries to NCBI PubMed we extracted 3,000 of abstracts. Extraction of associations between names of protein, gene and microRNA allowed us to reveal their possible impact into the development of MDD. In particular, we identified genes of the dopaminergic and serotoninergic systems and the glucocorticoid receptor, that are associated with altered expression levels that may occur as a result of adverse circumstances (e. g. deprivation) in early childhood and thus may play a role in the development of depression in adults in stressful situation or in difficult life circumstances. Such changes of gene expression levels can represent the so-called vicious circle of pathology.

Additionally, we identified relationships between epigenetic mechanisms of estrogen receptor regulation that may be responsible for the development of mental illness. The factors that contribute to the development of major depressive also include a particular attachment type of the child, the ability to form object relationship that depend on a early interaction of a child with an adult. The factors that may impact on the development of MDD may also include comorbid disorders such as borderline personality disorder or anxiety-depressive disorders, which both may be accompanied by more severe symptomatology and should be differentiated from other psychiatric disorder.

Conclusion: We evaluated the potential mechanisms for the development of major depressive disorder using text mining approach. The revealed reasons include a combination of biological and psychological factors. In particular, we found some of the molecular mechanisms responsible for the development of MDD, including gene regulation of the dopaminergic system. The role of these genes in the development of MDD is discussed in earlier studies [3]. At the same time the identified relationships between estrogen and glucocorticoid receptor regulation and the development of depression in adults is interesting for further investigation. The relationships between psychological and hereditary factors in the dynamics in patients with major depressive disorder are also of interest.

Funding: The study is supported by Russian Science Foundation grant No. 24-25-00453.

Список литературы/References

1. Uher R. at al. Major depressive disorder in DSM-5: implications for clinical practice and research of changes from DSM-IV. *Depress Anxiety.* 2014;31(6):459-471. doi 10.1002/da.22217
2. Thumfart K.M. et al. Epigenetics of childhood trauma: Long term sequelae and potential for treatment. *Neurosci Biobehav Rev.* 2022;132:1049-1066. doi 10.1016/j.neubiorev.2021
3. Tarasova O.A. et al. Chemical named entity recognition in the texts of scientific publications using the naïve Bayes classifier approach. *J Cheminform*. 2022;14(1):55. doi 10.1186/s13321-022-00633-4
4. Tarasova O.A. et al. Extraction of data on parent compounds and their metabolites from texts of scientific abstracts. *J Chem Inf Model.* 2021;61(4):1683-1690. doi 10.1021/acs.jcim.0c01054
5. Weber L. et al. HunFlair: an easy-to-use tool for state-of-the-art biomedical named entity recognition. *Bioinformatics*. 2021;37(17):2792-2794. doi 10.1093/bioinformatics/btab042

doi 10.18699/bgrs2024-7.3-06

Поиск мишеней для терапии болезни Паркинсона на основе анализа транскриптома черной субстанции модельных животных мышей с МФТП-индуцированным паркинсонизмом и дисфункцией глюкоцереброзидазы

Усенко Т.С.[1]*, Безрукова А.И.[1, 2], Башарова К.С.[1], Руденок М.М.[3], Шадрина М.И.[3], Сломинский П.А.[3], Пчелина С.Н.[1, 2]

[1] *Петербургский институт ядерной физики им. Б.П. Константинова НИЦ «Курчатовский институт», Гатчина, Россия*

[2] *Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, Санкт-Петербург, Россия*

[3] *Институт молекулярной генетики НИЦ «Курчатовский институт», Москва, Россия*

* *usenko_ts@pnpi.nrcki.ru*

Ключевые слова: болезнь Паркинсона; терапия; модельные животные; транскриптом

Мотивация и цель: Болезнь Паркинсона (БП) – распространенное нейродегенеративное заболевание, которое характеризуется гибелью дофаминергических нейронов в черной субстанции (ЧС) головного мозга [1]. Мутации в гене *GBA1*, кодирующем лизосомный фермент глюкоцереброзидазу (GCase), являются самым распространенным фактором генетического риска БП. БП, ассоциированная с мутациями в гене *GBA1* (GBA1-БП), является наиболее распространенной формой заболевания с известной этиологией [2, 3]. Цель данной работы заключалась в поиске потенциальных биомаркеров и терапевтических мишеней БП при дисфункции GCase на основе анализа транскриптома ЧС модельных животных.

Методы и алгоритмы: В ходе данной работы были сформированы 4 группы мышей самцов линии C57BL/6 по 4 животных в каждой группе со следующими инъекциями: 0.9 % раствора NaCl (контроль), нейротоксин 1-метил-4-фенил-1,2,3,6-тетрагидропиридина (MPTP) для индукции паркинсонизма, растворенный в 0.9 % физиологическом растворе и введенный подкожно в дозе 12 мг/кг дважды с интервалом 2 часа, селективный ковалентный ингибитор GCase кондуритол-β-эпоксид (CBE) для дисфункции GCase, растворенный в 0.9 % физиологическом растворе и введенный внутрибрюшинно разово в концентрации 100 мг/кг и одновременным введением CBE и MPTP (CBE/MPTP) в концентрациях, описанных выше для индукции паркинсонизма с дисфункцией GCase. В результате РНК-секвенирования дифференциально экспрессирующиеся гены (ДЭГ) в ЧС модельных животных определяли с помощью пакета DESeq2 (v3.12) по следующим критериям: кратность изменения экспрессии |Fold Change|<1.5 и p-value < 0.01 в среде R с использованием встроенного пакета DESeq2. Оценка экспрессии генов проводилась методом количественной ПЦР в режиме реального времени в мононуклеарах периферической крови, выделенных методом градиентного центрифугирования на градиенте фиколла, 14 пациентов с GBA1-БП, 30 пациентов с БП, 15 неврологически здоровых бессимптомных носителей мутаций в гене *GBA1* и 45 индивидуумов контрольной группы.

Результаты: Анализ дифференциальной экспрессии генов в клетках ЧС головного мозга мышей выявил 64 ДЭГ в группе MPTP, 2 ДЭГ в группе CBE, 37 ДЭГ в группе MPTP/CBE по сравнению с контролем, а также 23 ДЭГ в группе MPTP по сравнению с CBE, 23 ДЭГ MPTP/CBE по сравнению с CBE, 20 ДЭГ в группе MPTP/CBE по сравнению с MPTP. С помощью базы Gene Ontology было показано, что выявленные в группе MPTP/CBE по сравнению с MPTP и контролем ДЭГ вовлечены в такие биологические процессы, как нейрогенез, везикулярный транспорт и аутофагия, посредством которой деградирует белок альфа-синуклеин. Интересно отметить, что продукты генов с наиболее значимым изменением уровня экспрессии в группе MPTP/CBE по сравнению с группой CBE и контролем вовлечены в сигнальный путь PI3K/AKT/mTOR, являющийся ключевым звеном регуляции аутофагии. На основе результатов анализа транскриптома была оценена экспрессия генов, вовлеченных в PI3K/AKT/mTOR, а именно генов, вовлеченных в mTOR-зависимую аутофагию (*mTOR*, *MAP1LC3B*, *BECN1*, *SQSTM1*, *CTSD*). Показано увеличение экспрессии гена *MAP1LC3B* у пациентов с GBA1-БП и GBA1-носителей по сравнению с пациентами с сБП ($p < 0.05$), а также увеличение экспрессии гена *MAP1LC3B* у пациентов с GBA1-БП по сравнению с контролем ($p < 0.01$). Уровень экспрессии гена mTOR был повышен у пациентов с GBA1-БП и сБП по сравнению с GBA1-носителями и контролем ($p < 0.05$). Также было выявлено снижение экспрессии гена *SQSTM1* у пациентов с GBA1-БП по сравнению с пациентами с сБП ($p < 0.05$).
Выводы: Наши результаты позволяют предположить, что БП с дисфункцией GCase характеризуется выраженным нарушением пути PI3K/AKT/mTOR. Данный путь может рассматриваться как потенциальная терапевтическая мишень для БП, ассоциированной с мутациями в гене *GBA1*.
Финансирование: Исследование поддержано грантом Российского научного фонда (№ 24-15-00177).

Targets for therapy of Parkinson's disease based on transcriptome analysis of the substance nigra of a mouse model with MPTP-induced parkinsonism and glucocerebrosidase dysfunction

Usenko T.S.[1, 2]*, Bezrukova A.I.[1, 2], Basharova K.S.[1, 2], Rudenok M.I.[3], Shadrina M.I.[3], Slominsky P.A.[3], Pchelina S.N.[1, 2]

[1] *Petersburg Nuclear Physics Institute named by B.P. Konstantinov of National Research Center "Kurchatov Institute", Gatchina, Russia*

[2] *Pavlov First Saint-Petersburg State Medical University, St. Petersburg, Russia*

[3] *Institute of Molecular Genetics of National Research Center "Kurchatov Institute", Moscow, Russia*

* *usenko_ts@pnpi.nrcki.ru*

Key words: Parkinson's disease; therapy; model animals; transcriptome

Motivation and Aim: Parkinson's disease (PD) is a common neurodegenerative disease characterized by death of dopaminergic neurons in the substantia nigra (SN) of the brain [1]. Mutations in the *GBA1* gene, which encodes the lysosomal enzyme

glucocerebrosidase (GCase), are the most common genetic risk factor for PD. PD associated with mutations in the *GBA1* gene (GBA1-PD) is one of the most common form of disease with a widespread etiology [2, 3]. The goal of the work was to search for biomarkers and therapeutic targets for PD with GCase dysfunction based on transcriptome analysis of SN of animal models.

Methods and Algorithms: In the current study, 4 groups of male mice of the C57BL/6 line were formed, 4 animals in each group, following injections: 0.9 % NaCl solution (control), neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) for induction of parkinsonism, dissolved in a 0.9 % serial solution and administered subcutaneously at a dose of 12 mg/kg multiple times at 2 intervals, the selective covalent GCase inhibitor conduritol-β-epoxide (CBE) for GCase dysfunctions, dissolved in a 0.9 % serial solution and administered intraperitoneally in a single dose institutions 100 mg/kg and simultaneous administration of CBE and MPTP (CBE/MPTP) at the concentrations described above to induce parkinsonism with GCase dysfunction. Whole-transcriptome analysis is carried out in SN cells of the mouse brain. Differentially expressed genes (DEGs) are performed using the DESeq2 package (v3.12) according to the following criteria: fold change in expression |Fold Change|<1.5 and *p*-value<0.01 in R using DESeq2 package. Gene expression was assessed by quantitative real-time PCR in peripheral blood mononuclear cells isolated by Ficoll gradient centrifugation from 14 patients with GBA1-PD, 30 patients with PD, 15 neurologically healthy asymptomatic carriers of mutations in the GBA1 gene (GBA1-carries) and 45 individuals in the control group.

Results: Analysis of differential gene expression in SN cells of the mouse brain revealed 64 DEGs in the MPTP group, 2 DEGs in the CBE group, 37 DEGs in the MPTP/CBE group compared with the control, as well as 23 DEGs in the MPTP group compared with CBE, 23 MPTP DEGs /CBE vs. CBE, 20 DEGs in MPTP/CBE vs. MPTP group. Using the Gene Ontology database, it was shown that the DEGs identified in the MPTP/CBE group compared to MPTP and control are involved in biological processes such as neurogenesis, vesicular transport and autophagy, through which the protein alpha-synuclein is degraded. It is interesting to note that the gene products with the most significant changes in expression levels in MPTP/CBE group compared to CBE group and control are involved in PI3K/AKT/mTOR signaling pathway, which is a key link in the regulation of autophagy. Based on the results of our transcriptome analysis, the expression level of genes involved in PI3K/AKT/mTOR, namely genes involved in mTOR-dependent autophagy (*mTOR*, *MAP1LC3B*, *BECN1*, *SQSTM1*, *CTSD*), was assessed. An increase in the *MAP1LC3B* gene expression was shown in patients with GBA1-PD and GBA1-carriers compared to patients with PD ($p < 0.05$), as well as an increase in the *MAP1LC3B* gene expression in patients with GBA1-PD compared to controls ($p < 0.01$). The level of mTOR gene expression was increased in patients with GBA1-PD and PD compared to GBA1-carriers and controls ($p < 0.05$). A decrease in the *SQSTM1* gene expression was also found in patients with GBA1-PD compared to patients with PD ($p < 0.05$).

Conclusions: Thus, it can be assumed that PD with GCase dysfunction is characterized by disruption of the PI3K/AKT/mTOR pathway. This pathway can be considered as a potential therapeutic target for PD, associated with mutations in the *GBA1* gene.

Funding: The study is supported Russian Science Foundation (No. 24-15-00177).

Список литературы/References

1. Lill C.M. Genetics of Parkinson’s disease. *Mol Cell Probes*. 2016;30(6):386-396. doi 10.1016/j.mcp.2016.11.001
2. Neumann J., Bras J., Deas E. et al. Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. *Brain.* 2009;132(Pt. 7):1783-1794. doi 10.1093/brain/awp044
3. Emelyanov A.K., Usenko T.S., Tesson C., Senkevich K.A.. et al. Mutation analysis of Parkinson’s disease genes in a Russian data set. *Neurobiol Aging*. 2018;71:267.e7-267.e10. doi 10.1016/j.neurobiolaging.2018.06.027

doi 10.18699/bgrs2024-7.3-07

Expression of arginine vasopressin and corticoliberin receptor genes in the prefrontal cortex and hippocampus of mice with anxiety-like behavior phenotype

Belova V.*, Tuchina O.
Immanuel Kant Baltic Federal University, Kaliningrad, Russia
** belovaleria91@mail.ru*

Key words: PTSD; SPS; vasopressin; corticoliberin; mice; receptors; microglia

Motivation and Aim: Stress can disrupt normal brain processes and lead to the development of severe mental illnesses such as post-traumatic stress disorder (PTSD) [1]. The Single Prolonged Stress (SPS) protocol is one of the most commonly used rodent behavioral models employed by researchers to study PTSD [2]. This study focuses on how SPS affects changes in the expression of stress hormone receptors vasopressin (AVP) and corticotropin-releasing hormone (CRH) in the prefrontal cortex and hippocampus—brain structures responsible for the formation of fear and the consolidation of memory related to traumatic events [3]. Additionally, the level of neuroinflammation was assessed by counting hippocampal microglial cells responsible for the immune response in the body.
Methods and Algorithms: For the experiment, male wild-type mice were selected and divided into a control group and a stress group. The experiment involved 6 animals in the control group and 5 animals in the stress group. The animals were housed in a vivarium at Immanuel Kant Baltic Federal University (IKBFU). Experimental animals in the stress group were subjected to the Single Prolonged Stress (SPS) protocol, which comprised acute and chronic phases, for one month. No additional mental stress was applied to the animals in the control group during this period. At the conclusion of the experiment, behavioral phenotyping was conducted using the "Open Field" and "Elevated Plus Maze" tests. Subsequently, animals from both the control and stress groups underwent decapitation and perfusion. RNA was extracted from the prefrontal cortex and hippocampus of the left hemisphere using ExtractRNA kit (Evrogen, Russia), followed by reverse transcription. Gene expression levels were assessed using real-time PCR (BioRad C-1000). The right hemisphere was utilized for immunohistochemical examination. Serial sections with a thickness of 50 microns were obtained using a cryotome (Cryostat Microtome KD-3000, KEDEE, China). Immunohistochemical staining was performed using the Iba-1 microglial marker (ab178847, Abcam, United Kingdom) with a secondary antibody (ab150073, Abcam, United Kingdom). Iba+ cells in three zones of the hippocampus (CA1, CA2/3, and DG) were manually counted at 20x magnification using a Leica DM4000 B LED fluorescence microscope (LEICA Microsystems, Germany) equipped with LAS V.4.3 software.
Results: In the "Open Field" test, mice from the stress group exhibited significantly reduced motor activity and spent more time in the periphery compared to mice from the control group. In the "Elevated Plus Maze" test, males from the stress group spent more time in closed arms than in open arms, unlike males from the control group. Consequently, mice from the stress group displayed decreased exploratory behavior and

elevated levels of anxiety. In the prefrontal cortex of males from the stress group, an increase in the expression levels of arginine-vasopressin receptors Avpr1a and Avpr1b was observed, while the expression of the Avpr2 receptor and the corticotropin-releasing hormone receptor Crhr1 remained unchanged compared to the control group. In the hippocampus of males from the stress group, there was a decrease in the expression levels of Avpr1b and Crhr1 receptors, but no changes in the expression levels of Avpr1a and Avpr2 were observed. The number of Iba+ microglial cells was increased in the stress group compared to the control group in the CA1, CA2/3, and DG regions of the hippocampus, indicating the presence of neuroinflammation in mice from the stress group.

Conclusion: Therefore, SPS contributes to a decline in exploratory behavior and an increase in anxiety levels in mice from the stress group. Additionally, SPS induces neuroinflammation in the hippocampus and also has varying effects on the expression profiles of vasopressin and corticotropin-releasing hormone receptors in the prefrontal cortex and hippocampus of mice.

Funding: The study was carried out with the financial support by the Institute of Medicine and Life Sciences of the Immanuel Kant Baltic Federal University.

References

1. Morris M.C., Compas B.E., Garber J. Relations among posttraumatic stress disorder, comorbid major depression, and HPA function: a systematic review and meta-analysis. *Clin Psychol Rev*. 2012;32(4):301-315. doi 10.1016/j.cpr.2012.02.002
2. Keller S.M. et al. Inhibiting corticosterone synthesis during fear memory formation exacerbates cued fear extinction memory deficits within the single prolonged stress model. *Behav Brain Res*. 2015;287:182-186. doi 10.1016/j.bbr.2015.03.043
3. Etkin A., Wager T.D. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am J Psychiatry*. 2007;164(10):1476-1488. doi 10.1176/appi.ajp.2007.07030504

doi 10.18699/bgrs2024-7.3-08

Involvement of cholinergic system in the induction of some early genes in hippocampal cells after stress depends on the novelty of stress context

Dobryakova Y.V.*, Deryabina A.K., Fedulova A.A., Koryagina A.A., Bolshakov A.P.
Institute of Higher Nervious Activity and Neurophysiology, RAS, Moscow, Russia
** julkadobr@gmail.com*

Key words: stress; acetylcholine; scopolamine; hippocampus; early genes

Motivation and Aim: The stress response is a universal response of the body to various changes in the external environment. The stress response includes many components, among which are nonspecific components such as the activation of the sympatho-adrenal axis and the hypothalamic-pituitary-adrenal axis, As well as specific components that involve the activation of various brain regions and signaling systems depending on the characteristics of the stressor. Changes in the expression of a number of genes are often considered as a result of the activation of cells in a particular brain region, including early genes such as egr1, fos, npas4, and others. The induction of early gene expression is a complex process, and the involvement of various signaling systems in the induction of diverse genes remains a poorly researched area.

Studies have shown that stress activates various signaling systems in the brain, including the noradrenergic, serotonergic, cholinergic, and other systems. It has been demonstrated that the expression of a number of early genes depends on the release of noradrenaline. The involvement of the cholinergic system in regulating the transcriptional response during the development of the stress response remains virtually unexplored, so the aim of our work was to analyze the involvement of metabotropic muscarinic receptors in the regulation of gene transcription in the acute period after brief unavoidable swimming stress.

Methods and Algorithms: Our study included two series. In the first series, Wistar rats were handled and then subjected to unavoidable swimming stress in a cylinder for the Porsolt test for 15 minutes. This series included 4 groups of animals. (1) Control animals injected with saline; (2) control animals injected with scopolamine (antagonist of muscarinic receptors); (3) animals injected with saline and then subjected to swimming; (4) animals injected with scopolamine and then subjected to swimming. Each group contained 7–8 animals.

In second series, Wistar rats were handled and then, for 4 days, daily were subjected to swimming stress for 5 min. On the fifth day, some animals were subjected to swimming for 15 min as in the first series. The animal groups in this series were as follows: (1) intact animals not subjected to training for 4 days; (2) control animals injected with saline on the fifth day; (3) animals injected with scopolamine on the fifth day; (4) animals injected with saline and then to swimming on the fifth day; (5) animals injected with scopolamine and then to swimming on the fifth day.

In both series, the animals were decapitated 45 min after the onset of stress, the brain was extracted, and the dorsal and ventral hippocampi were isolated for subsequent analysis of gene expression using quantitative PCR.

Results: We found that, in the first series of experiments, swimming induced an increase in the expression of fos and npas4 genes in both hippocampal parts and egr1 only in the ventral hippocampus. Scopolamine suppressed stress-induced increase in the expression of npas4 and egr1 but not fos gene.
In the second series of experiments, stress caused by swimming on the fifth day also induced an increase in fos expression in both hippocampal parts and an increase in egr1 expression in the ventral hippocampus. Like in the first series, scopolamine did not affect fos expression after stress; however, in contrast to the first series, scopolamine has no effect on egr1 expression in the ventral hippocampus.
Conclusion: The acute unavoidable stress induces expression of early genes in both dorsal and ventral hippocampi; however, inhibition muscarinic receptors suppressed stress-induced changes in the gene expression only in the case of presentation of novel stress context. Presumably, involvement of cholinergic system in the induction of stress-associated changes in the gene expression is determined by novelty of stress context.
Funding: The study was performed in the frameworks of the State task of the Institute of Higher Nervous Activity and Neurophysiology RAS.

doi 10.18699/bgrs2024-7.3-09

The impact of stress on pro-inflammatory cytokines and glial markers expression in rat strains with different excitability levels

Korolevich D.[1]*, Kunafin D.[1], Vylegzhanina A.[1], Sidorova M.[1], Shalaginova I.[1], Eresko S.[2], Airapetov M.[2], Pavlova M.[3], Zachepilo T.[3], Dyuzhikova N.[3]

[1] *Immanuel Kant Baltic Federal University, Kaliningrad, Russia*

[2] *Institute of Experimental Medicine, Saint Petersburg, Russia*

[3] *Pavlov Institute of Physiology, Russian Academy of Sciences, Saint Petersburg, Russia*

* *korolevich.dana@gmail.com*

Key words: stress; neuroinflammation; rat; astrocytes; cytokines

Motivation and Aim: Stress is a critical factor in the development of post-stress disorders, yet its underlying mechanisms remain unclear. Research indicates that stress is associated with neuroinflammation, which plays a significant role in anxiety and depression. This study focuses on the pro-inflammatory cytokines Interleukin 1 beta (IL-1β) and tumor necrosis factor (TNF), produced during pathological conditions and stress responses [1]. GFAP and Iba1 are glial markers of astrocytes and microglia respectively. This research aims to assess changes in pro-inflammatory cytokines and glial markers in the brain of rats with different genetically determined levels of nervous system excitability at different periods after stress exposure.

Methods and Algorithms: The experiment was carried out on male five-month-old rats of two strains: with a high threshold (HT) of nervous system excitability (low excitable) and a low threshold (LT) of nervous system excitability (highly excitable). Each experiment involved 24 animals, 6 in each group. The animals belong to the biocollection of the “Pavlov Institute of Physiology" (RAS) [2]. Experimental animals were subjected to psychogenic long-term emotional and pain stress in accordance with the Gecht protocol: every day for 15 days the animals were exposed to 6 unsupported (10 seconds each) and 6 reinforced current (2.5 mA, 2 ms) light signals with a probability of 50 %. Animals of the control and experimental groups were decapitated 7 and 24 days after the end of stress. For immunohistochemical studies, additional perfusion was performed. Total RNA was isolated from the amygdala, hippocampus and prefrontal cortex of the right hemisphere (extractRNA, Evrogen, Russia), reverse transcription was performed, and the mRNA level was assessed using RT-PCR (BioRad C-1000). The corresponding structures of the left hemisphere were used to assess protein levels (SEA563Ra, SEA133Ra, SEC288Ra kits, Cloud-Clone Corp., China) by ELISA. For immunohistochemical study, the left hemisphere was taken, the animal brain was cut on a cryotome (Cryostat Microtome KD-3000, KEDEE, China) in 50 μm thick slices and subjected to immunohistochemical staining for astrocytic marker (PAA068Ra02, Cloud-Clone Corp., China) with secondary antibody (A-21207, Thermo Fisher Scientific, USA). Data were analyzed under a Leica DM4000 B LED fluorescence microscope (LEICA Microsystems, Germany) with LAS V.4.3 software. Cells were counted at a magnification of 20×. Three regions of interest each were analyzed in the prefrontal cortex (IL, PrL, Cg1) at 2.70, 2.20, 1.70 from bregma and in the hippocampus (CA1,

CA2/3 and DG) at –2.30, –3.30, –4.80 from bregma [3]. In the amygdala, cells were counted without distinguishing individual regions at levels –2.12 to –2.56 of bregma. Cells were counted in 3 fields of view in each region of interest. Astrocytic cell counts were calculated on 8 slices from each animal using the ImageJ analysis program.

Results: Motor activity decreased in both strains after 24 days, and exploratory activity in the Elevated Plus Maze test also decreased; however, in the LT strain, these changes were evident in a wider range of measured parameters. TNF expression increases in the amygdala in LT strain rats on day 24 after stress at both the mRNA and protein levels. In the hippocampus, TNF at the mRNA level is significantly higher in the stressed group of the HT strain 24 days after stress compared with the control. The level of IL-1β protein significantly increased only in the amygdala of stressed LT rats 24 days after stress, while mRNA expression remained at the control level. Iba1 expression increases in the amygdala in low excitable HT rats 7 days after stress at the mRNA level, while no changes are observed at the protein level. The number of GFAP+ cells decreases in the amygdala, CA1, CA3 and DG of hippocampus and in prefrontal cortex in LT rats 7 days after stress. In HT rats, at the same time point after stress, the number of cells decreases only in the CA1 area of hippocampus.

Conclusion: Thus, in rats with high and low hereditary excitability of the nervous system, different levels of expression of pro-inflammatory and glial markers and their specific changes under the influence of stress were revealed. In LT rats, these changes were evident in a wider range of measured parameters, indicating a heightened sensitivity to stress which is in line with their more pronounced behavioral response to stress. However, in HT rats, specific changes such as increased TNF mRNA in the hippocampus and altered Iba1 expression in the amygdala suggest that they are also significantly affected by stress. These unique patterns observed in HT rats indicate a need for more detailed studies on the long-term consequences of stress in this strain. Further research could elucidate the delayed effects of stress and the potential for developing chronic conditions, providing insights into tailored interventions for individuals with different genetic predispositions.

Funding: The study is supported by funds from the program strategic academic leadership “Priority 2030” IKBFU I. Kant and funds from the federal budget within the framework of the state assignment of the Federal State Budgetary Institution Pavlov Institute of Physiology (No. 1021062411629-7-3.1.4).

References

1. Munshi S. et al. Repeated stress induces a pro-inflammatory state, increases amygdala neuronal and microglial activation, and causes anxiety in adult male rats. *Brain Behav Immun*. 2020;84:180-199. doi 10.1016/j.bbi.2019.11.023
2. Vaido A.I., Shiryaeva N.V., Pavlova M.B. et al. Selected lines of rats with high and low thresholds of excitability: a model for studying maladaptive states dependent on the level of excitability of the nervous system. *Lab Anim Sci Res*. 2018;3:12-22. doi 10.29296/2618723X-2018-03-02 (in Russian)
3. Paxinos G., Watson C. The rat brain in stereotaxic coordinates: hard cover edition. Elsevier, 2006

doi 10.18699/bgrs2024-7.3-11

Phospholipid supplementation inhibits social odor discrimination in mice

Morozova M.[1], Andrejeva J.[2]*, Snytnikova O.[3], Boldyreva L.[1], Tsentalovich Yu.[3], Kozhevnikova E.[2, 4]

[1] *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

[2] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[3] *International Tomography Center, SB RAS, Novosibirsk, Russia*

[4] *Novosibirsk State Agrarian University, Novosibirsk, Russia*

* *andreeva.e@mcb.nsc.ru*

Key words: phospholipids; diet; metabolism; odor discrimination; social behavior

Motivation and Aim: The dietary phospholipids (PLs) are the promising supplements are that are commonly found as natural food ingredients and as emulsifier additives. There is a broad consensus that supports the beneficial impact of dietary PL supplementation on human health and the management of various diseases, including neurological disorders. The present study aimed at evaluating the effect of major PLs found as food supplements on social behavior in mice. Here, the effect of short-term high dietary PLs content was studied in terms of social odor discrimination and social interactions with male and female intruders in male mice.

Methods: The open field test is utilized to evaluate both the locomotor and exploratory behavior of the test animals, and it can also measure the overall anxiety levels. The odor preference assay enables the assessment of an animal's attraction/aversion to different scents without direct contact with the source of the odor. This experiment aimed at exploring aspects of social behavior in mice, such as social interaction, aggression, and mating. Brain metabolomic analysis was performed to assess changes in intermediary metabolism and neurotransmitter biosynthesis.

Results: PLs-fed male mice tend to lose preference toward female odor and fail to discriminate socially relevant odors. At the same time, test animals recognize non-social odors. We also found that PLs affected the social behavior of the test males, rendering them indiscriminative of male and female intruders during direct contact. Brain metabolomic profiling revealed no major changes in the intermediary metabolism and neurotransmitter biosynthesis. At the same time, intranasal PLs application resembled the same effects as dietary supplementation.

Conclusion: The findings presented here might be important to understand the physiological relevance of PLs as modulators of signal transduction in sensory organs and other systems. Here we provide in vivo data that could inspire studies on the molecular interactions of PLs with transmembrane receptors and second messenger signal transducers and reevaluate their role in ligand sensing.

Funding: This research and the APC was funded by the Russian Science Foundation, grant number 20-74-10022-П.

doi 10.18699/bgrs2024-7.3-12

Serotonin receptors interactions in the regulation of behavior and central nervous system functioning

Naumenko V.S.*, Kondaurova E.M., Ilchibaeva T.V., Tsybko A.S., Rodnyy A.Ya.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** naumenko2002@mail.ru*

Key words: serotonin receptors; receptor heterodimerization; brain; behaviour

Motivation and Aim: Brain serotonin (5-HT) is a principal player in the mechanisms of brain and behavioral plasticity. This neurotransmitter plays an important role in the regulation of multiple physiological functions and types of behavior. Polyfunctionality of the brain 5-HT is due to a large variety of the receptors mediating 5-HT action on neurons. Currently, 14 different subtypes of 5-HT receptors are known in 5-HT receptor superfamily. With exception of the 5-HT_3 receptor, that is ligand-gated ion channel, all other 5-HT receptors belong to the G-protein coupled receptor family [1]. Considerable structure-specific differences in the density as well as in physiological effects of 5-HT receptors was shown [1, 2]. The recent data indicate that all receptor types have multiple signaling cascades and some of the receptors could couple with more than one G-protein. Additionally, some of receptors could form hererodimers with other brain receptors and proteins that significantly affect central nervous system functioning. Such interplay between different kinds of 5-HT receptors multiplies functional effects of 5-HT system [3–6].
Methods and Algorithms: Our studies showed that heterodimeric complexes could be formed both by different receptors of the same brain system and by receptors of different brain systems even if receptors belong to different superfamilies. Thus, the ability to form heterodimers was shown for 5-HT1A and 5-HT7 receptors from the serotonin G-protein-coupled receptors family as well as for serotonin G-protein-coupled receptor 5-HT2A and brain-derived neurotrophic factor TrkB receptor, belonging to the family of receptors with enzymatic activity. It was demonstrated that formation of heterodimeric complexes significantly affects the function of at least one of composing receptors that sometimes has an impressive behavioral effects. Additionally, it was revealed that serotonin G-protein-coupled 5-HT4 receptor is able to form heterodimers with molecule of cell adhesion L1. This interaction significantly affects neuronal architecture.
Conclusion: Growing body of evidence indicate that receptors heterodimerization is an additional factor in the regulation of the functional activity of the brain neurotransmitter and neurotrophic systems and, hence, behavior. The obtained data draw attention to receptor complexes as new targets for pharmacological correction of behavioral pathologies.
Funding: The cost of animal maintenance was supported by the basic research project # FWNR-2022-0023; the study was supported by the Russian Science Foundation (grant No. 22-15-00011).

References

1. Barnes N.M., Sharp T. A review of central 5-HT receptors and their function. *Neuropharmacology.* 1999;38(8):1083-152
2. Saudou F., Hen R. 5-Hydroxytryptamine receptor subtypes: molecular and functional diversity. *Adv Pharmacol.* 1994;30:327-380
3. Kulikov A.V., Gainetdinov R.R., Ponimaskin E., Kalueff A.V., Naumenko V.S., Popova N.K. Interplay between the key proteins of serotonin system in SSRI antidepressants efficacy. *Expert Opin Ther Targets.* 2018;22(4):319-330
4. Popova N.K., Naumenko V.S. 5-HT1A receptor as a key player in the brain 5-HT system. *Rev Neurosci.* 2013;24(2):191-204
5. Naumenko V.S., Popova N.K., Lacivita E., Leopoldo M., Ponimaskin E.G. Interplay between serotonin 5-HT1A and 5-HT7 receptors in depressive disorders. *CNS Neurosci Ther*. 2014;20(7):582-590
6. Naumenko V.S., Bazovkina D.V., Kondaurova E.M. On the functional cross-talk between Brain 5-HT1A and 5-HT2A receptors. *Zhurnal Vysshei Nervnoi Deiatelnosti imeni I.P. Pavlova.* 2015;65(2):240-247

doi 10.18699/bgrs2024-7.3-13

Wireless magnetoelectric interface for deep brain stimulation

Romashchenko A.V.[1, 2]*, Chernozem R.V.[2], Surmenev R.A.[2], Moshkin M.P.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *National Research Tomsk Polytechnic University, Tomsk, Russia*

* *aromash2006@gmail.com*

Key words: core-shell nanoparticles, magnetoelectric effect, nose-to-brain transport, deep brain stimulation

Motivation and Aim: Magnetoelectric (ME) nanoparticles (NPs) can transform magnetic-field energy to electric-field energy and vice versa, which makes them attractive for the development of novel approaches to wireless control of biological systems. ME NPs have a range of potential applications, such as drug delivery, cancer treatment, control over cell differentiation and stimulation of neurons.[1-7] The wireless brain–machine interface is one of the most intriguing and promising issues that can be addressed using ME NPs. Nonetheless, to provide an effective particle–neuron coupling for in vivo deep brain stimulation, ME NPs must meet several requirements: low toxicity, high efficiency of ME coupling at weak magnetic fields, controlled delivery.

Methods and Algorithms: A new approach for ultrafast *in situ* fabrication of stabilised colloidal highly efficient magnetoelectric (ME) core–shell nanoparticles (NPs) was developed on the basis of biocompatible magnetic $MnFe_2O_4$ (MFO) and ferroelectric $Ba_{0.85}Ca_{0.15}Zr_{0.1}Ti_{0.9}O_3$ (BCZT) using a microwave-assisted hydrothermal method. MFO NPs were used as a control for all our experiments to evaluate the role of mechanical stimulation in observed biological effects of MFO@BCZT NPs due to possible magnetic particle rotation in AC magnetic field. The toxicity of the synthesised NPs with and without AMF (7 mT, 10 Hz) exposure was evaluated by the MTT assay. We performed T_2*-weighted MRI and TEM to assess the ability of MFO@BCZT NPs to be taken up by the mouse olfactory epithelium and transported throughout the olfactory system. The ability of MFO@BCZT NPs to alter olfactory neuronal activity *in vivo* during AMF stimulation was studied by Mn-enhanced MRI. Manganese ions act as agonists to potential-dependent calcium channels, and hence their accumulation is proportional to the neuronal functioning. The mice were then exposed to clean air or odour (0.01% acetophenone) with or without AMF stimulation (7 mT, 10 Hz). Throughout all the experiments, the animals were kept in their respective home cages.

Results: Opposite effects on growth of cancer and normal cells were revealed *in vitro* by electrostimulation via ME MFO@BCZT NPs using a low-intensity AC magnetic field (7 mT, 10 Hz). Enhanced uptake and trans-synaptic axonal transport of MFO@BCZT NPs to the brain and were detected *in vivo*. Using T2*-weighted MRI and specific inhibitors of axonal transport we showed the prevalence of transneuronal movement of MFO@BCZT NPs from mouse nasal cavity to the brain.

The application of the AC magnetic field did not influence the MOB T_1-weighted MRI signal during clean air or odor presentation in control and MFO NP–treated animals but enhanced this signal in MFO@BCZT NP–treated mice.

The behavioural experiments indicated that mice were able to detect and recognise odour stimuli with equal proficiency in the presence or absence of NPs or AC magnetic field. This ability is determined by the number of olfactory neurons and the activity of MOB

interneurons.[8] There was statistical significance only for the effect of the AC magnetic field on the total time of sniffing of the odorant in MFO@BCZT NP–treated, but not MFO NP–treated mice. This parameter directly depends on the activity or number of olfactory receptor neurons.[8]

Conclusion: Thus, in this study, a novel advanced strategy is established based on non-invasive intracellular ME nanosized electrodes for neurological theranostics and targeted drug delivery to the central nervous system

by means of ME MFO@BCZT NPs as a non-invasive platform for wireless electrical stimulation.

Funding: The study is supported by RSF grant № 23-14-00179.

References

1. Rodzinski A. et al.Targeted and controlled anticancer drug delivery and release with magnetoelectric nanoparticles. *Sci. Rep*. 2016;6(1):20867
2. Stewart T.S. et al. Magnetoelectric nanoparticles for delivery of antitumor peptides into glioblastoma cells by magnetic fields. *Nanomed.* 2018;13(4):423-438
3. Smith I.T. et al. Nanomedicine and nanobiotechnology applications of magnetoelectric nanoparticles. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* 2023;15(2):1849
4. Betal S. et al. Magneto-elasto-electroporation (MEEP): In-vitro visualization and numerical characteristics. *Sci. Rep.* 2016;6(1):32019
5. Bok I. et al. In silico assessment of electrophysiological neuronal recordings mediated by magnetoelectric nanoparticles. *Sci. Rep.* 2022;12(1):8386
6. Chernozem R. V. et al. Novel biocompatible magnetoelectric MnFe2O4 Core@ BCZT shell nano–hetero-structures with efficient catalytic performance. *Small* 2023;2302808
7. Zhang Y. et al. Magnetoelectric nanoparticles incorporated biomimetic matrix for wireless electrical stimulation and nerve regeneration. *Adv. Health. Mat.* 2021;10(16):2100695
8. Kikuta S. et al. Longer latency of sensory response to intravenous odor injection predicts olfactory neural disorder. *Sci. Rep*. 2016;6(1):35361

doi 10.18699/bgrs2024-7.3-14

Differential gene expression and neuroimmune responses in the amygdala of rat strains with high and low nervous system excitability

Shalaginova I.[1]*, Pavlova M.[2], Zachepilo T.[2], Dyuzhikova N.[2]

[1] *Immanuel Kant Baltic Federal University, Kaliningrad, Russia*

[2] *Pavlov Institute of Physiology, RAS, St. Petersburg, Russia*

* *shalaginova_i@mail.ru*

Key words: transcriptome analysis; rats; excitability; neuroinflammation; amygdala

Motivation and Aim: The aim of this study is to investigate the transcriptome profiles and neuroimmune reactions to stress in the amygdala of rat strains with contrast levels of nervous system excitability. Understanding the molecular and immune mechanisms associated with these differences can provide insights into the genetic and neurobiological factors that may influence excitability, stress responses, and related behavior abnormalities. Previous research has highlighted the importance of genetic predisposition in excitability to severity of neuroinflammation and post-stress changes in physiology. However, comprehensive transcriptomic data specific to the amygdala, particularly in the context of excitability level and neuroimmune interactions, remain unclear. This study seeks to fill this gap by providing gene expression profiles and identifying potential regulatory pathways involved in the differences observed between these rat strains.

Methods and Algorithms: The experiment was conducted on five-month-old male rats from two strains: high threshold (HT) of nervous system excitability (low excitability) and low threshold (LT) of nervous system excitability (high excitability). Each experiment included 24 animals, with 6 in each group in case of rtPCR, ELISA and IHC analysis and 5 in case of RNAseq. The animals were part of the biocollection at the "Pavlov Institute of Physiology" (RAS). Experimental animals expose to long-term emotional and pain stress using the Gecht protocol and control and experimental groups were decapitated 7 and 24 days post-stress. Total RNA was isolated from the right hemisphere's amygdala (extractRNA, Evrogen, Russia), followed by RT-PCR (BioRad C-1000) to assess mRNA levels of IL1b, TNF, IL6, IL10, BDNF, IBA1, GFAP). Protein levels in the left hemisphere were measured using ELISA kits for of IL1b, TNF, IL6, IL10 and IBA1 (SEA563Ra, SEA133Ra, SEC288Ra, Cloud-Clone Corp., China). The left hemisphere was also used for immunohistochemical staining for an astrocytic marker GFAP in amygdala (PAA068Ra02, Cloud-Clone Corp., China).

To analyze the transcriptomes of the amygdala in two rat strains tissue samples were collected and total RNA was extracted from tissues using Trisol reagent and the PureLink RNA micro Kit (Invitrogen). Sequencing was was performed at the Genoanalitika Lab, Moscow, Russia. Libraries for sequencing were prepared using the NEBNext® Ultra™ II RNA Library Prep Kit (NEB). Sequencing was done on a HiSeq1500 (Illumina), generating at least 20 million paired-end 50-nucleotide reads per sample. Read mapping and counting were performed with STAR 2.7.9a. Genome: Rnor_6.0. Annotation: Ensembl v.99. Differential expression: Deseq2 v.1.28.1. Genes

with Padj < 0.05 and |log2FC|>0,38 were designated as differentially expressed genes (DEGs). To identify overrepresented Gene Ontology (GO) terms in a list of DEGs, the DEGs were analyzed for functional enrichment using the DAVID tool.

Results: The RNA-seq analysis revealed significant differences in gene expression between the intact rats of two strains. We identified 152 upregulated and 105 down-regulated DEGs in the amygdala of LT strain compared to the HT strain. Top upregulated genes in LT strain vs HT strain: AABR07041096.1 (lincRNA), Serpinf2, Nit2, Pyroxd2. Pyroxd2 is predicted to enable oxidoreductase activity and is involved in mitochondrial organization. It is located in the mitochondrial matrix. So, upregulation of this gene in the LT rat strain's amygdala suggests increased mitochondrial activity, which could be a response to heightened cellular stress and energy demands. Enhanced mitochondrial function might be necessary to support the increased excitability and metabolic needs of neurons, potentially influencing neuroinflammatory processes and stress-related behaviors. Pdilt – Protein Disulfide Isomerase-like is involved in protein folding and acts as a chaperone. It is located in the endoplasmic reticulum (ER) lumen and is active in germ cell migration and spermatid development. its function in the brain has not been studied, but there is evidence of the expression of this gene in the cortex and pituitary gland of rats [1]. The upregulation of Pdilt in the LT rat strain's amygdala could indicate an increased demand for protein folding and ER stress response. Proper protein folding is crucial for maintaining cellular function, especially under stress conditions. Among the downregulated genes can be noted Ift88, that crucial for intracellular transport processes. Its downregulation could disrupt cellular trafficking and signaling, possibly affecting cellular responses to environmental stressors; Rexo4 is involved in DNA binding, nuclease activity, DNA catabolic process, and DNA repair. Downregulation of Rexo4 may lead to impaired DNA repair mechanisms, potentially increasing genomic instability [2] and vulnerability to neuroinflammation and cellular stress. Alb facilitates fatty acid, modified amino acid, and zinc ion binding. Involved in the regulation of sleep. Reduced Alb levels may affect the transport and availability of essential fatty acids and ions, potentially disrupting neuronal function.

Functional Enrichment: The Gene Ontology (GO) analysis indicates that the DEGs are significantly involved in biological processes (BP) such as oxygen transport, angiogenesis, lipid metabolism and, most importantly, in immunity and innate immunity; cellular components (CC) such as the extracellular matrix and endoplasmic reticulum; and molecular functions (MF) such as oxidoreductase and glycosyltransferase activities.

PCR and ELISA analyses found that in the amygdala of highly excitable rats (LT), stress leads to increased levels of the pro-inflammatory cytokines IL1b and TNF compared to controls. This indicates a pronounced neuroinflammatory response in these rats under stress. Additionally, immunohistochemistry data showed a decrease in the number of GFAP-expressing astrocytes in the amygdala of high excitability rats (LT) under stress. Which is consistent with data on the loss of glial cells in some brain structures in response to stress [3].

Conclusion: The RNA-seq results revealed key differences in gene expression between the LT and HT strains. In LT rats, the upregulation of genes associated with inflammatory and metabolic processes suggests an inherent predisposition to heightened stress responses and neuroinflammation. This aligns with the observed increases in IL1b and TNF levels under stress, indicating that these rats have a molecular profile primed for robust inflammatory responses.

Funding: The study is supported by funds from the program strategic academic leadership “Priority 2030” IKBFU I. Kant and funds from the federal budget within the framework of the state assignment of the Federal State Budgetary Institution Pavlov Institute of Physiology (No. 1021062411629-7-3.1.4).

References

1. Darbellay F., Necsulea A. Comparative transcriptomics analyses across species, organs, and developmental stages reveal functionally constrained lncRNAs. *Mol Biol Evol.* 2020;37(1):240-259. doi 10.1093/molbev/msz212
2. Shcherbinina V., Pavlova M., Daev E., Dyuzhikova N. Rats selected for different nervous excitability: long-term emotional–painful stress affects the dynamics of DNA damage in cells of several brain areas. *Int J Mol Sci*. 2024;25(2):994. doi 10.3390/ijms25020994
3. Fang H., Bing X., Lili W. Loss of glial cells of the hippocampus in a rat model of post-traumatic stress disorder. *Neurochem Res*. 2015;40(5):942-951. doi 10.1007/S11064-015-1549-6

doi 10.18699/bgrs2024-7.3-15

Investigation of the epigenetic effects of IVF on circadian rhythms of descendents

Silvanovich E.[1, 3]*, Zuev D.S.[1, 2], Sharapova M.B.[1], Romashchenko A.V.[1], Moshkin M.P.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *National Research University ITMO, St. Petersburg, Russia*

* *e.silvanovich@alumni.nsu.ru*

Key words: in-vitro fertilization; IVF; behavior; RNA sequencing; circadian rhythm; sleep

Motivation and Aim: During observations, performed in our laboratory, it has been noticed, that IVF descendents have a higher body mass than the control mice. After a sequence of behavioral experiments it was stated that the only parameter significantly differing among groups is the day-night distribution of mice activity: IVF group showed a higher activity rate during the light time, in which animals usually sleep. Thus, the aim of further nervous tissue analysis was to indicated, whether there are any epigenetic alterations in genes expression that can possibly provide this shift of activity and, if such, what molecular pathways are involved in it.

Materials and Methods: Samples for transcriptomic analysis were taken from the frontal cortex of both hemispheres of the mouse brain. Read quality assessment was performed using the FastQC tool [1], followed by trimming using Trimmomatic [2]. Genomic indexing was conducted using the Kallisto package [3] based on the Mus musculus Ensemble Mg39 reference genome. Subsequent analysis of the counted reads was performed using the Phantasus program [4] – PCA analysis was conducted, and differential expression analysis was performed using the built-in Limma program [5]. Enrichment of the results of differential expression analysis was carried out using the Enrichr database [6] and GSEA MSigDB [7].

Results: Subsequently, transcriptomic analysis was performed on 36 mouse brain samples – 8 samples from the control group and 28 from the test group. Fifteen thousand genes were identified with non-zero readings and were subsequently used in the analysis. Principal component analysis (PCA) revealed the presence of two distinct clusters separated by 1 component (1st component 19 %, 2nd component 10 %). Analysis of differential gene expression identified genes that significantly increased their expression in the test group of animals. Enrichment analysis of the 250 most highly expressed genes showed significant overlap with REACTOME database attributes, involved in circadian regulation and mitochondria functioning- REACTOME_CIRCADIAN_CLOCK, REACTOME_MITOCHONDRIAL_BIOGENESIS and REACTOME_TRANSCRIPTIONAL_ACTIVATION_OF_MITOCHONDRIA_BIOGENESIS. Especially, significantly increased expression was observed among genes-key players in the CLOCK complex, which is the main regulator of circadian rhythms.

Conclusions: Thus, we have demonstrated that in animals conceived through IVF, changes in circadian rhythm of metabolism are associated with alterations in the genetic regulation of this process.

Funding: This work was supported by grant RFBR No. 23-14-00179.

References

1. Lo C.C., Chain P.S.G. Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics*. 2014;15:366
2. Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120
3. https://github.com/pachterlab/kallisto
4. Kleverov M. et al. Phantasus: web-application for visual and interactive gene expression analysis. *bioRxiv*. 2022
5. Ritchie M.E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47-e47
6. https://github.com/wjawaid/enrichR
7. https://github.com/GSEA-MSigDB

doi 10.18699/bgrs2024-7.3-16

Effect of *Tnf* gene knockout on behavior, BDNF expression, and brain serotonin system during long-term administration of dexamethasone

Skotnikova A., Adonina S., Moskaliuk V., Kulikova E., Bazovkina D.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** annakonstantinovna2000@gmail.com*

Key words: dexamethasone; TNF; serotonin; behavior; BDNF

Motivation and Aim: The pro-inflammatory cytokine tumor necrosis factor (TNF) is one of the key mediators between the immune and central nervous systems. TNF is involved in the regulation of behavior and the level of BDNF (brain-derived neurotrophic factor, a key neurotrophin within the brain), and the functioning of the serotonin system of the brain. In turn, chronic administration of the synthetic glucocorticoid dexamethasone can lead to changes in brain BDNF levels and behavioral impairments. Thus, the aim of this work was to study the effect of complete knockout of the *Tnf* gene on the sensitivity of mice to long-term administration of dexamethasone.
Methods and Algorithms: Adult C57BL/6J mice (WT) and mice with *Tnf* gene knockout (TNF KO) were injected daily with DEX (4 mg/kg, i. p.) or saline. The behavior of the animals was assessed in the following paradigms: the open field (OF), the tail suspension test (TST), the elevated plus maze (EPM) and the novel object (NO) tests.The level of expression of genes encoding BDNF, its TrkB and p75 receptors and key elements of brain serotonin system in several brain structures (prefrontal cortex, hippocampus, midbrain) were measured by real-time RT-PCR. The levels of BDNF and its precursor proBDNF proteins were measured by Western blot analysis, and the levels of serotonin (5-HT) and its metabolite 5-HIAA in the same brain structures were examined by HPLC. The data were analyzed by two-way ANOVA followed by LSD post-hoc test.
Results: Long-term administration of dexamethasone led to decrease in motor activity in the OF test ($p < 0.05$), an increase in depressive-like behavior in the TST ($p < 0.05$) and in anxiety-like behavior in the EPM test ($p < 0.05$) only in TNF KO mice, but not in WT mice. In contrast, deterioration in cognitive performance in the NO test was observed only in WT mice ($p < 0.05$). An increase in the expression of the *Bdnf* gene at the mRNA level in hippocampus ($p < 0.05$) was observed only in TNF KO mice ; the increase in the proBDNF protein level in the frontal cortex was revealed in both strains of mice ($p < 0.05$ for WT and $p < 0.001$ for TNF KO). Dexamethasone also led to a decrease in the expression of the gene encoding the 5-HT7 receptor ($p < 0.05$) in the midbrain of TNF KO animals.
Conclusion: Thus, the depressive-like and anxiety-like behavior of TNF KO mice caused by long-term administration of dexamethasone can be explained by an increase in the level of proBDNF and the proBDNF/BDNF ratio in the hippocampus, detected only in animals of this strain. This, in turn, may be associated with differences in neurogenesis in the hippocampus in knockout animals. The data obtained allow us to expand our understanding of the influence of the immune system on neurotrophic support and the functioning of the serotonin system of the brain.

doi 10.18699/bgrs2024-7.3-17

The role of BDNF in the mechanisms of autistic-like behavior in BTBR mice

Tsybko A.[1], Ilchibaeva T.[1], Shcherbakova A.[1, 2], Kaminskaya Y.[1], Eremin D.[1], Naumenko V.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *antoncybko@mail.ru*

Key words: brain-derived neurotrophic factor (BDNF); autism; autism spectrum disorders; BTBR mice

Motivation and Aim: The mechanisms underlying autism spectrum disorder (ASD) are still poorly understood, but impaired neuroplasticity undoubtedly plays an important role in the development of the disease. Brain-derived neurotrophic factor (BDNF) significantly implicated in regulation of neuronal plasticity and behavior. In addition, a large number of human and animal studies indicate the involvement of BDNF in the pathogenesis of ASD [1, 2]. The reduced BDNF mRNA and protein levels in the hippocampus and the frontal cortex of BTBR mice (validated as the model for ASD) have been demonstrated in a few studies [3, 4] but these data are scarce and incomplete.
Methods and Algorithms: Here we investigated *Bdnf* exon transcripts expression as well as BDNF and its precursor proBDNF proteins in the hippocampus (HC), frontal cortex (FC), striatum (ST) and midbrain (MB) of BTBR mice in comparison with normosocial C57Bl/6 mice. Compensation of the deficits of mature BDNF in BTBR mice was performed by either i.c.v. injection of recombinant BDNF protein or its overexpression in neurons using AAV vectors encoding *Bdnf* gene under synapsin promoter.
Results: It was found that mRNA levels of *Bdnf* exons 1–4 were decreased in HC of BTBR mice. In the ST of BTBR mice the number of transcripts *Bdnf* exons 1, 2, 4 was also decreased. The mRNA levels of *Bdnf* exons 1 and 6 were decreased in FC of BTBR mice. The proBDNF level as well as proBDNF/BDNF ratio were increased in HC and ST of BTBR mice.
Administration of BDNF protein failed to affect the behavior of BTBR mice. However, a decrease in the proBDNF/BDNF ratio in ST was detected after i.c.v injection of BDNF. Hippocampal overexpression of BDNF significantly reduced anxiety-related and stereotyped behavior of BTBR mice without affecting their social interest. In contrast, overexpression of BDNF in the FC of BTBR mice exclusively increased social interest but not other types of behaviors.
Conclusion: Thus, we have shown for the first time the significant changes in expression of specific *Bdnf* transcripts that may underlie impaired BDNF maturation in BTBR mice as evidenced by the proBDNF prevalence. Also, it was demonstrated that direct compensation of mature BDNF deficit by induction of BDNF overexpression can ameliorate autistic-like behavior in BTBR mice, but this effect is strongly dependent on the target brain structure.
Funding: The study is supported by RSF grant (No. 22-15-00028).

References

1. Liu S.H., Shi X.J., Fan F.C., Cheng Y. Peripheral blood neurotrophic factor levels in children with autism spectrum disorder: a meta-analysis. *Sci Rep*. 2021;11:15. doi 10.1038/s41598-020-79080-w
2. Reim D., Schmeisser M.J. Neurotrophic factors in mouse models of autism spectrum disorder: Focus on BDNF and IGF-1. *Adv Anat Embryol Cell Biol*. 2017;224:121-134. doi 10.1007/978-3-319-52498-6_7
3. Stephenson D.T., O'Neill S.M., Narayan S. et al. Histopathologic characterization of the BTBR mouse model of autistic-like behavior reveals selective changes in neurodevelopmental proteins and adult hippocampal neurogenesis. *Mol Autism*. 2011;2:7. doi 10.1186/2040-2392-2-7
4. Daimon C.M., Jasien J.M., Wood W.H. et al. Hippocampal transcriptomic and proteomic alterations in the BTBR mouse model of autism spectrum disorder. *Front Physiol*. 2015;6:324. doi 10.3389/fphys.2015.00324

doi 10.18699/bgrs2024-7.3-18

Effects of IVF on mice offspring behavior and metabolism

Zuev D.S.[1, 2]*, Anudarieva A.A.[2], Romashchenko A.V.[1], Moshkin M.P.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *zuevdaniil.zuevdaniil@gmail.com*

Key words: in-vitro fertilization; IVF; behavior; metabolism; respiratory exchange coefficient; RER; circadian rhythm; sleep

Motivation and Aim: *In-vitro* fertilization (IVF) has become a common assistive reproductive technology, however, it poses a number of risks, including increased incidence of obstetric complications and developmental disorders [1]. It is imperative to study the effects of IVF, not only to improve the outcomes of the procedure, but to gain more insight into the human embryogenesis in general. One of the less severe IVF effects on health, shown on large cohorts, is increased obesity, surprisingly without a significant increase in diabetes incidence [2]. In our laboratory we have observed similar effect in IVF-derived mice, that had higher body mass, but showed no signs of glucose intolerance. We hypothesized that increased body mass may be arise from differences in feeding behavior, locomotion or respiratory substrate utilization. In this study we assess whether differences between IVF-derived and naturally conceived CD1 mice are associated with behavioral differences.

Methods and Algorithms: Study included 2 groups of CD1 male mice: naturally conceived (control group, Ctrl, n=8) and IVF-derived (IVF group, n=10). IVF was conducted as previously described [3]. Mice were kept at artificial lighting conditions 10D:14L, lights-off event was considered a zeitgeber event (0:00 in zeitgeber time, ZT). At 10–11 weeks of age, mice were weighted, glucose tolerance test was conducted, and subsequently their behavior was recorded in the PhenoMaster cage PhenoMaster (TSE Systems, Germany). Recordings included locomotor activity, prolonged periods of inactivity (sleep), food and water consumption, indirect calorimetry (O_2 & CO_2 concentrations). PhenoMaster additionally calculates respiratory exchange ratio (CO_2 exhaled divided by O_2 consumed), an important metabolic parameter, that shows what substrate (fats or carbohydrates) the animal is oxidizing at the moment. We used Mann–Whitney U-test (two-sided) for all the group comparisons.

Results: IVF mice demonstrated significantly higher body mass than controls ($p < 0.01$) while areas under curve in glucose tolerance test were not significantly different. Surprisingly, average travelled distance of IVF mice was significantly higher than one of controls ($p < 0.05$), while food consumption per unit body mass was not significantly different. Average O_2 consumption and respiratory exchange ratio (RER) did not show any significant differences between groups. To better capture differences, we have averaged all behavioral time series across 4 days of experiment, to obtain a 24-hour long time series for each individual animal. We observed that the IVF group had a significant increase in RER at the end of the light phase (20–22 ZT, $p < 0.05$). Food consumption was also significantly higher around that period (hours 20, 23 ZT, $p < 0.05$). Travelled distance was significantly higher in IVF mice in many timepoints throughout the 24 hours. Daily sleep time was significantly lower in IVF group, compared to controls

both during light phase ($p < 0.001$), when mice typically have an a daily minimum of activity, and during dark phase ($p < 0.01$), when mice exhibit peak activity.
Conclusion: The data above demonstrates that IVF-derived mice tend to be more active and sleep less than naturally conceived ones. They also tend to have elevated respiratory exchange ratio during the light phase of the day, which is a sign of decreased lipid catabolism and increased glucose catabolism [4]. This suggests that IVF-derived mice higher body mass may be a result of differences in sleep-wake cycles.
Funding: The study is supported by the Russian Science Foundation (grant No. 23-14-00179).

References

1. von Wolff M., Haaf T. In vitro fertilization technology and child health. *Dtsch Arztebl Int*. 2020;117(3):23-30. doi 10.3238/arztebl.2020.0023
2. Norrman E., Petzold M., Gissler M. et al. Cardiovascular disease, obesity, and type 2 diabetes in children born after assisted reproductive technology: A population-based cohort study. *PLoS Med*. 2021;18(9):e1003723. doi 10.1371/journal.pmed.1003723
3. Anisimova M.V., Gon Y., Kontsevaya G.V. et al. Body composition as an indicator of metabolic changes in mice obtained by in vitro fertilization. *Vavilovskii Zhurnal Genetiki i Selektsii.* 2023;27(4):357-365. doi 10.18699/VJGB-23-43
4. Ono-Moore K.D., Rutkowsky J.M., Pearson N.A. et al. Coupling of energy intake and energy expenditure across a temperature spectrum: impact of diet-induced obesity in mice. *Am J Physiol Endocrinol Metab*. 2020;319(3):E472-E484. doi 10.1152/ajpendo.00041.2020

8 Симпозиум «Микробиология и биотехнологии: компьютерные и экспериментальные подходы»

Symposium “Microbiology and biotechnologies: computational and experimental approaches”

doi 10.18699/bgrs2024-8.1-01

Биосенсоры на основе люциферазы для выявления активности шаперонов и отслеживания биоэнергетики митохондрий

Аль Ибрахим Р.Н.[1]*, Алексеева М.Г.[2], Баженов С.В.[1], Фомин В.В.[1, 4], Кессених А.Г.[4], Мавлетова Д.А.[2], Нестеров А.А.[2, 3], Полуэктова Е.Ю.[2], Даниленко В.Н.[2], Манухов И.В.[1]

[1] *Московский физико-технический институт, Долгопрудный, МО, Россия*

[2] *Института общей генетики им. Н.И. Вавилова РАН, Москва, Россия*

[3] *Институт инженерной экологии РУДН, Москва, Россия*

[4] *Лаборатория микробиологии Росбиотех, Москва, Россия*

* *rahaf.alebrahim6@mail.ru*

Ключевые слова: биосенсоры; шапероны; митохондрии; болезнь Паркинсона

В этом исследовании мы создали набор конструкций с генами люциферазы из различных источников, таких как *Luciola mingrelica*, *Photorhabdus luminescens* и *Aliivibrio fischeri*, для экспрессии в бактериальных и эукариотических клетках. Реципиентные клетки с этими конструкциями позволяют анализировать внутриклеточную активность шаперонов, оценивать пул восстановленных эквивалентов и проводить токсикологические исследования.

Пара генов люцифераз из *A. fischeri* и *P. luminescens*, различающихся по термостабильности (оптимум люминесценции при 25 и 38 °C соответственно), применялись для исследования дезагрегазной активности штамма *Limosilactobacillus fermentum* U-21. Этот штамм секретирует шапероны в среду и рассматривается как кандидат для разработки лекарств, известных как дезагрегазы, для терапии болезни Паркинсона [1]. Мы показали, что секретируемый белок, кодируемый локусом C0965_000195, согласно анализу последовательности и структуры, является ClpL. С помощью вышеописанной пары люцифераз показано, что ClpL в гетерологичной системе, в клетках *Escherichia coli*, может компенсировать дефицит гена *clpB* и улучшать рефолдинг белков. Эксперименты *in vitro* продемонстрировали, что культуральная среда, содержащая секретируемые клетками штамма U-21 белки, включая ClpL, может предотвращать термоденатурацию люцифераз. Эти результаты позволяют предположить, что ClpL, обладающий свойствами дезагрегазы, может значительно способствовать фармбиотическим свойствам *L. fermentum* U-21, что делает его перспективным кандидатом для разработки терапевтических стратегий против нейродегенеративных заболеваний.

Гены светлячковой люциферазы *L. mingrelica* и бактериальной из *P. luminescens*, оптимизированной для экспрессии в млекопитающих были успешно трансфецированы в клетки HEK293t. Конструкции с геном *luc L. mingrelica* для экспрессии люциферазы как в цитоплазме, так и в митохондриальном матриксе позволили анализировать внутриклеточный пул АТФ и влияние работы дыхательной цепи митохондрий на него. Конструкции с генами *luxAB P. luminescens* позволяют анализировать внутриклеточный пул восстановленного ФМН. Оксидоредуктаза LuxG дополнительно позволяет оценивать пул НАДН в

клетке. Локализация в цитоплазме и матриксе митохондрий бактериальной и светлячковой люцифераз позволяет оценивать влияние различных биологически активных веществ на энергетичесике процессы в разных компартментах клетки.
Финансирование: Конструирование биосенсоров и исследование шаперонной активности поддержано РНФ 22-14-00124.
Клонирование гена clpL и работа с L. fermentum U-21 проведены в ИОГен РАН за счёт ГЗ 122022600163-7.
Влияние биологически активных веществ на биосенсоры исследовано при поддержке Минобрнауки РФ, проект FSMF-2023-0010.

Luciferases based biosensors for detecting chaperone activity and monitoring mitochondrial bioenergetics

Al Ebrahim R.N.[1]*, Alekseeva M.G.[2], Bazhenov S.V.[1], Fomin V.V.[1, 4], Kessenikh A.G.[4], Mavletova D.A.[2], Nesterov A.A.[2, 3], Poluektova E.U.[2], Danilenko V.N.[2], Manukhov I.V.[1]

[1] *Moscow Institute of Physics and Technology, Dolgoprudny, Moscow region, Russia*
[2] *Laboratory of Genetics of Microorganisms, Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, Russia*
[3] *Institute of Environmental Engineering, RUDN University, Moscow, Russia*
[4] *Laboratory of Microbiology BIOTECH University, Moscow, Russia*
* *rahaf.alebrahim6@mail.ru*

Key words: biosensors; chaperones; mitochondria; Parkinson's disease

In this study, we developed a series of constructs with luciferase genes from various sources, such as *Luciola mingrelica*, *Photorhabdus luminescens*, and *Aliivibrio fischeri*, for expression in bacterial and eukaryotic cells. Recipient cells with these constructs allow for the analysis of intracellular chaperone activity, assessment of the reduced equivalents pool, and toxicological studies.
A pair of luciferase genes from *A. fischeri* and *P. luminescens*, differing in thermostability (optimal luminescence at 25 and 38 °C respectively), were used to investigate the disaggregase activity of the *Limosilactobacillus fermentum* U-21 strain. This strain secretes chaperones into the medium and is considered a candidate for the development of disaggregase drugs for Parkinson's disease therapy [1]. We showed that the secreted protein encoded by the C0965_000195 locus, according to sequence and structural analysis, is ClpL. Using the above-mentioned pair of luciferases, it was shown that ClpL in a heterologous system, in Escherichia coli cells, can compensate for the clpB gene deficiency and improve protein refolding. *In vitro* experiments demonstrated that the culture medium containing proteins secreted by the U-21 strain, including ClpL, can prevent the thermal denaturation of luciferases. These results suggest that ClpL, possessing disaggregase properties, can significantly contribute to the pharmabiotic properties of *L. fermentum* U-21, making it a promising candidate for the development of therapeutic strategies against neurodegenerative diseases.
The luciferase genes from the firefly *L. mingrelica* and the bacterium *P. luminescens*, optimized for expression in mammals, were successfully transfected into HEK 293T cells. Constructs with the *luc* gene from *L. mingrelica* for luciferase expression in both

the cytoplasm and mitochondrial matrix allowed for the analysis of the intracellular ATP pool and the impact of mitochondrial respiratory chain activity on it. Constructs with the *luxAB* genes from *P. luminescens* enable the analysis of the intracellular reduced FMN pool. The *LuxG* oxidoreductase further allows for the assessment of the NADH pool in the cell. Localization of bacterial and firefly luciferases in the cytoplasm and mitochondrial matrix makes it possible to evaluate the effects of various bioactive substances on the energy processes in different cell compartments.

Funding: The construction of biosensors and the investigation of chaperone activity were supported by the Russian Science Foundation (grant number 22-14-00124).

Cloning of the *clpL* gene and work with *L. fermentum* U-21 were carried out at the Vavilov Institute of General Genetics Russian Academy of Sciences, funded by State Assignment 122022600163-7.

The study of the effects of bioactive substances on biosensors was supported by the Ministry of Science and Higher Education of the Russian Federation, project FSMF-2023-0010.

Список литературы/References

1. Stavrovskaya A.V., Voronkov D.N., Marsova M.V. et al. Effects of the pharmabiotic U-21 in a combined neuroinflammatory model 476 of Parkinson's disease in Rats. *Bull Exp Biol Med*. 2024;177(2):193-199. doi 10.47056/0365-9615-2024-177-2-193-199

doi 10.18699/bgrs2024-8.1-02

Стратегические перспективы создания микробиом направленных продуктов в России

Даниленко В.Н.
Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия
valerid@vigg.ru

Ключевые слова: фармабиотики; постбиотики; микробиом (микробиота)

Мотивация и цель: Успехи и достижения науки о жизни в последние десятилетия позволили сделать прорыв в создании лекарственных препаратов по различным направлениям, в том числе в области онкологии, кардиологии и других. Вместе с тем, существующими подходами не удается решить проблему множественной лекарственную устойчивости и создания новых эффективных антибактериальных препаратов, создания препаратов нового поколения для лечения неврологических и нейродепрессивных состояний. Требуются новые подходы для создания препаратов, снимающих побочные действия при лучевой, химиотерапии и иммунотерапии, онкозаболеваний. Природоподобные технологии, открывают новые горизонты создания лекарственных средств нового поколения, адаптированных к организму человека. Микробиота кишечника человека и животных является резервуаром полезных бактерий, функциональных генов, сигнальных молекул и метаболитов, определяющих позитивный гомеостаз организма, и служащая их источником для создания новых биотерапевтических препаратов различной направленности.
Методы и алгоритмы: В ИОГен РАН разработан алгоритм поиска генов и бактерий, их содержащих, с заданными свойствами. Разработаны методы экспрессии скрининга среди коллекции штаммов с заданными свойствами. Для отобранных кандидатов препараты проводится изучение их свойств на адекватных моделях культур клеток, грызунов.
Результаты: В лаборатории генетики микроорганизмов ИОГен РАН проводятся работы по созданию фармабиотиков и постбиотиков на основе штаммов: Levilactobacillus brevis 47f для снятия побочных эффектов мукозитной природы при лучевой терапии, Limosilactobacillus fermentum U21 для комбинированного лечения пневмонии, Bifidobacterium longum GT15 и Bifidobacterium adolescentis в комбинированной терапии депрессивных состояний. Для каждого из предложенных кандидатов в препараты осуществлён большой комплекс исследований, в том числе основанных на омиксных и геномных технологиях, а также определенный цикл доклинических исследований по их эффективности и безопасности.
Заключение: Сегодня микробиота кишечника является источником фармабиотиков и постбиотиков (метабиотиков) с нейромодулирующими,

иммуномодулирующими и противовоспалительными свойствами с установленными механизмами действия, биологически активными ингредиентами, их определяющими. В качестве источника препаратов следует рассматривать не только микробиом здоровых людей различного возраста и регионов, но и животных, включая диких и домашних животных, пчелу и т. д. Конечная цель – создание препаратов фармабиотиков и ингредиентов для продуктов питания, для лечения и профилактики заболеваний различной этиологии.

Финансирование: Работа частично выполнена в рамках гранта Министерства науки и высшего образования Российской Федерации на реализацию масштабных научных проектов по приоритетным направлениям развития науки и технологий (проект № 075-15-2024-638).

Strategic perspectives on the development of microbiome-targeting products in Russia

Danilenko V.N.

N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia
valerid@vigg.ru

Key words: pharmabiotics; postbiotics; microbiome (microbiota)

Motivation and Objective: Advances and achievements in life sciences over the past decades have led to breakthroughs in the creation of pharmaceutical products across various fields, including oncology, cardiology, and others. However, existing approaches have not resolved the issues of multiple drug resistance and the development of new, effective antibacterial drugs, nor have they succeeded in creating a new generation of drugs for treating neurological and neurodepressive conditions. New approaches are required to develop drugs that mitigate side effects during radiation, chemotherapy, and immunotherapy for cancer treatment. Biomimetic technologies open new horizons for creating next-generation drugs that are adapted to the human body. The gut microbiota of humans and animals is a reservoir of beneficial bacteria, functional genes, signaling molecules, and metabolites that determine positive organism homeostasis and serve as a source for developing new biotherapeutic drugs with various targets.

Methods and Algorithms: The VIGG RAS has developed an algorithm for identifying genes and the bacteria containing them with specified properties. Methods for expression screening among a collection of strains with desired properties have been developed. For selected candidate drugs, their properties are studied on appropriate models, including cell cultures and rodents.

Results: The Laboratory of Microorganism Genetics at VIGG RAS is working on the development of pharmabiotics and postbiotics based on strains such as L. brevis 47f to mitigate mucositis-related side effects during radiation therapy, L.

fermentum U21 for combined pneumonia treatment, B. longum GT15 and B. adolescentis for combined therapy of depressive states. A comprehensive set of studies has been conducted for each of the proposed drug candidates, including omics and genomic technologies, along with a series of preclinical studies on their efficacy and safety.

Conclusion: Today, the gut microbiota is a source of pharmabiotics and postbiotics (metabiotics) with neuromodulatory, immunomodulatory, and anti-inflammatory properties with established mechanisms of action and biologically active ingredients. The microbiomes of healthy individuals of various ages and regions, as well as animals, including wild and domestic animals and bees, should be considered sources for these drugs. The ultimate goal is to develop pharmabiotic drugs and ingredients for food products for the treatment and prevention of diseases of various etiologies.

Funding: This work was partially supported by a grant from the Ministry of Science and Higher Education of the Russian Federation for the implementation of large-scale scientific projects in priority areas of scientific and technological development (project No. 075-15-2024-638).

Список литературы/References

1. Averina O.V., Poluektova E.U., Zorkina Y.A., Kovtun A.S., Danilenko V.N. Human Gut Microbiota for Diagnosis and Treatment of Depression. *Int J Mol Sci*. 2024;25:5782. doi 10.3390/ijms25115782
2. Odorskaya M.V., Mavletova D.A., Nesterov A.A. et al. The use of omics technologies in creating LBP and postbiotics based on the *Limosilactobacillus fermentum* U-21. *Front Microbiol*. 2024;15:1416688. doi 10.3389/fmicb.2024.1416688
3. Stavrovskaya A.V., Voronkov D.N., Marsova M.V., Olshansky A.S., Gushchina1 A.S., Danilenko V.N., Illarioshkin S.N. Effects of the pharmabiotic U-21 in a combined neuroinflammatory model of Parkinson's disease in rats. *Bull Exp Biol Med*. 2024;177(2):193-199. doi 10.47056/0365-9615-2024-177-2-193-199
4. Danilenko V.N., Devyatkin A.V., Marsova M.V., Shibilova M.U., Ilyasov R.A., Shmyrev V.I. Common Inflammatory Mechanisms in COVID-19 and Parkinson's Diseases: The Role of Microbiome, Pharmabiotics and Postbiotics in Their Prevention. *J Inflammation Res.* 2021;14:6349-6381. doi 10.2147/JIR.S333887
5. Yunes R.A., Poluektova E.U., Belkina T.V., Danilenko V.N. Lactobacilli: Legal Regulation and Prospects for New Generation Drugs. *Appl Biochem Microbiol*. 2022;58(5):652-664

doi 10.18699/bgrs2024-8.1-03

Предсказание биологических механизмов влияния кишечных микробов на результаты иммунотерапии меланомы с использованием метагеномных данных

Захаревич Н.В.[1], Морозов М.Д.[1], Канаева В.А.[1, 3], Иванов А.Б.[1, 2], Ульянцев В.И.[2], Климина К.М.[1], Олехнович Е.И.[1]*

[1] *ФГБУ ФНКЦ ФХМ им. Ю.М. Лопухина ФМБА России, Москва, Россия*

[2] *Университет ИТМО, Санкт-Петербург, Россия*

[3] *Московский физико-технический институт, Долгопрудный, Россия*

** jeniaole13@mail.ru*

Ключевые слова: микробиота кишечника; иммунотерапия рака; меланома; MAG

Мотивация и цель: Мировое научное сообщество проводит масштабные исследования, направленные на определение степени влияния микробиоты кишечника человека на результаты иммунотерапии злокачественных опухолей. Согласно полученным недавно результатам, характеристики микробиоты кишечника человека не только положительно влияют на результаты иммунотерапии, но и могут быть переданы [1–5]. Тот факт, что фенотип ответчика передается при переносе микробиоты, позволяет предположить, что этот феномен может быть связан с конкретной бактерией или набором бактерий (или какой-либо другой бактериальной характеристикой или метаболитом), которые могут быть выделены и использованы в качестве адъюванта для улучшения результатов иммунотерапевтического лечения. Однако, несмотря на большое количество опубликованных исследований, ученые до сих пор не пришли к единому мнению относительно микробных детерминант ответа на иммунотерапию меланомы. Ранее опубликованные метаанализы частично ответили на этот вопрос [6, 7], но полного понимания биологических процессов, лежащих в основе этого явления, пока нет. При этом идентификация бактерий-кандидатов осложняется локальностью собранных когорт, ограничениями по количеству собранных образцов, а также некоторыми другими объективными факторами. Недавно наша исследовательская группа выявила последовательные метагеномные биомаркеры кала, связанные с эффективностью иммунотерапии меланомы, используя общепринятые методы таксономической и функциональной аннотации [8]. Следующим шагом нашего исследования, было использование метагеномики с геномным разрешением, методов сравнительной геномики и метаболической реконструкции для изучения возможных механизмов участия микробиоты кишечника в ответе на терапию метастатической меланомы.

Методы и алгоритмы: В данном исследовании использовались методы биоинформатики, такие как геномно-разрешенная метагеномика, профилирование штаммов, сравнительная геномика и метаболическая реконструкция.

Результаты: Используя загруженные из базы данных NCBI/EBI 680 метагеномов кала из 7 опубликованных исследований пациентов с меланомой [1–5, 7, 9], мы собрали нередуцированный каталог бактериальных геномов (далее metagenome-assembled genomes – MAGs, геномы, собранные из метагеномов), который

включал 1422 MAG'a. Согласно качеству собранных геномов по критериям Genomic Standards Consortium [10], собранный каталог включал 1006 высококачественных и 416 MAG среднего качества. Согласно таксономической аннотации, из всего каталога 1416 являлись геномами бактерий и 6 – геномами архей. Список наиболее распространенных фил включал Firmicutes (902 генома), Actinobacteria (261 геном), Bacteroidetes (148 геномов), Proteobacteria (59 геномов), а также другие филумы (52 генома). После этого было проведено выявление биомаркеров, связанных с результатами иммунотерапии, осуществленное согласно ранее описанному протоколу [8]. В результате проведенного анализа был получен список из 137 MAGs-маркеров, отличающих пациентов по исходу иммунотерапии: 84 для положительного исхода иммунотерапии (R-группа, англ. responder) и 54 для отрицательного (NR-группа, англ. non-responder). В 6 и более исследованиях с положительным исходом иммунотерапии были связаны 5 MAG'ов включая *Bifidobacterium adolescentis*, *Bifidobacterium* unclassified, *Gemmiger qucibalis* и *Barnesiella intestinihominis*.

Следующим шагом нашей работы стало сравнение выявленных групп биомаркеров по функциональному содержанию в соответствии с классификациями KEGG и MetaCyc [11, 12]. В частности, мы обнаружили 41 группу ортологий по классификации KEGG и 63 по классификации MetaCyc, которые продемонстрировали значительные различия в представленности в группах MAG-биомаркеров. Примечательно, что все эти выявленные группы генов были повышены в R-группе. Результаты анализа обогащения группами генов показал, что 6 путей KEGG и 4 пути MetaCyc, включая связанные с биосинтезом аминокислот и кобаламина, были значительно повышены в R-группе. Кроме того, мы изучили связь между MAG-биомаркерами и вышеупомянутыми путями. Результаты этого анализа позволили выделить 5 родов (*Faecalibacterium*, *Blautia*, *Bacteroides*, *Bifidobacterium* и *Ruminococcus*), которые содержали наибольшее количество кластеров генов, относящихся к этим путям.

Используя методы метаболической реконструкции и функциональной аннотации, мы выявили сдвиг баланса прототрофии в сторону ауксотрофии аминокислот у пациентов с негативными результатами иммунотерапии. Кроме того, исследование путей биосинтеза короткоцепочечных жирных кислот, известных своей иммуномодулирующей ролью, выявило дифференциальное обилие этих путей среди конкретных MAG. Среди прочего, наличие кобаламин-зависимого пути синтеза ацетата Вуда–Люнгдаля было напрямую связано с позитивным ответом на иммунотерапию меланомы.

Выводы: В целом, полученные нами результаты позволили углубить понимание биологических механизмов влияния микробиома кишечника на результаты иммунотерапии меланомы и заложить основу для дальнейших исследований, направленных на повышение эффективности данного вида лечения с помощью модуляции микробиома.

Финансирование: Исследование поддержано грантом Российского Научного Фонда соглашение № 22-75-10029. Информация о проекте доступна по ссылке https://rscf.ru/project/22-75-10029/.

Predicting biological mechanisms of gut microbial influence on melanoma immunotherapy outcomes using metagenomic data

Zakharevich N.V.[1], Morozov M.D.[1], Kanaeva V.A.[1, 3], Ivanov A.B.[1, 2], Ulyantsev V.I.[2], Klimina K.M.[1], Olekhnovich E.I.[1]*

[1] *Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia*

[2] *ITMO University, Saint Petersburg, Russia*

[3] *Moscow Institute of Physics and Technology, Moscow, Russia*

* *jeniaole13@mail.ru*

Key words: gut microbiota; cancer immunotherapy; melanoma; MAG

Motivation and Aim: The global scientific community is conducting extensive research to determine the degree of influence of the human gut microbiota on the outcomes of immunotherapy for malignant tumors. Characteristics of the human gut microbiota not only positively influence the outcome of immunotherapy, but can also be transferred [1–5]. The fact that the responder phenotype is transferred by microbiota transfer suggests that this phenomenon may be associated with a specific bacterium or set of bacteria (or some bacterial characteristic or metabolite) that can be isolated and used as an adjuvant to improve the outcome of immunotherapeutic treatment. However, despite a large number of published studies, researchers still do not agree on the gut microbial determinants of response to melanoma immunotherapy. Previously published meta-studies have partially answered this question [6, 7], but a complete understanding of the biological processes underlying this phenomenon remains to be seen. However, the identification of candidate bacteria is complicated by the location and limitations of the samples collected, as well as some other objective factors. Recently, Olekhnovich and co-authors [8] identified consistent stool metagenomic biomarkers associated with melanoma immunotherapy efficacy using community-accepted methods of taxonomic and functional annotation. The next step of our study which we described here is to use enhanced methods including genome-resolved metagenomics, comparative genomics, and metabolic reconstruction methods to explore possible mechanisms of gut microbiota involvement in response to metastatic melanoma therapy.

Methods and Algorithms: This study used advanced bioinformatics techniques such as genome-resolved metagenomics, strain profiling, comparative genomics, and metabolic reconstruction to refine and develop the proposed concepts.

Results: Using our collected from NCBI/EBI database 680 stool metagenomes from 7 published studies from patients with melanoma [1–5, 7, 9], we assembled a non-redundant catalog of bacterial genomes (further metagenome-assembled genomes – MAGs) that included 1,422 genomes. According to the quality of the assembled genomes by Genomic Standards Consortium criteria [10], our genomes included 1,006 high-quality and 416 medium quality MAGs. According to the taxonomic annotation of the entire catalog, 1,416 were bacterial genomes of and 6 were archaeal. The list of the most common phyla includes Firmicutes (902 genomes), Actinobacteria (261 genomes), Bacteroidetes (148 genomes), Proteobacteria (59 genomes), and other

phylums (52 genomes). After that the biomarkers linked to immunotherapy outcome have been evaluated according to early described protocol [8]. As a result of the analysis, we obtained a list of 137 MAGs-biomarkers that distinguished patients by the immunotherapy outcome: 84 for positive immunotherapy outcome (R-group) as well as 54 for negative (NR-group). Consistent in 6 or more studies were 5 MAGs linked to the positive immunotherapy outcome including *Bifidobacterium adolescentis*, unclassified *Bifidobacterium*, *Gemmiger qucibalis*, and *Barnesiella intestinihominis*. The next step in our work was to compare the identified groups of biomarkers by functional content according to KEGG and MetaCyc classifications [11, 12].

Specific gene groups that distinguish functional categories among MAG biomarkers have been identified. Specifically, we found 41 KOG and 63 RXN categories that showed significant differences. Notably, all of these identified gene groups were upregulated in the R-group. The results of gene set enrichment analysis revealed that 6 KEGG pathways, and 4 MetaCyc pathways including associated with amino acid and cobalamin biosynthesis were significantly upregulated in the R-group. In addition, we explored the relationship between MAG-biomarkers and the aforementioned immunotherapy-relevant pathways. The results of this analysis, highlighted the top 5 genera that contained the highest number of gene groups from these pathways. These genera were *Faecalibacterium*, *Blautia*, *Bacteroides*, *Bifidobacterium*, and *Ruminococcus*.

Using metabolic reconstruction and functional annotation methods we identified a shift in the balance of amino acid prototrophy to auxotrophy in patients with negative immunotherapy outcomes. Furthermore, our investigation of the biosynthetic pathways of short-chain fatty acids, known for their immunomodulatory role, revealed a differential abundance of these pathways among the specific MAGs. Among others, the cobalamin-dependent Wood–Ljungdahl pathway of acetate synthesis was directly associated with responsiveness to melanoma immunotherapy.

Conclusion: In summary, our findings have advanced the understanding of the biological mechanisms of gut microbiome influence on melanoma immunotherapy outcome and provided a foundation for further investigations aimed at enhancing immunotherapy efficacy through microbiome modulation. In summary, our findings have advanced the understanding of the biological mechanisms of gut microbiome influence on melanoma immunotherapy outcome and provided a foundation for further investigations aimed at enhancing immunotherapy efficacy through microbiome modulation.

Funding: Financial support for this study was provided by the Russian Science Foundation under the grant number 22-75-10029, available at https://rscf.ru/project/22-75-10029/.

Список литературы/References

1. Frankel A.E., Coughlin L.A., Kim J. et al. Metagenomic shotgun sequencing and unbiased metabolomic profiling identify specific human gut microbiota and metabolites associated with immune checkpoint therapy efficacy in melanoma patients. *Neoplasia*. 2017;19(10):848-855
2. Gopalakrishnan V., Spencer C.N., Nezi L. et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science*. 2018;359(6371):97-103
3. Matson V., Fessler J., Bao R. et al. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science*. 2018;359(6371):104-108
4. Baruch E.N., Youngster I., Ben-Betzalel G. et al. Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients. *Science*. 2021;371(6529):602-609
5. Davar D., Dzutsev A.K., McCulloch J.A. et al. Fecal microbiota transplant overcomes resistance to anti-PD-1 therapy in melanoma patients. *Science*. 2021;371(6529):595-602

6. Limeta A, Ji B, Levin M, Gatto F, Nielsen J. Meta-analysis of the gut microbiota in predicting response to cancer immunotherapy in metastatic melanoma. *JCI Insight*. 2020;5(23):e140940
7. Lee K.A., Thomas A.M., Bolte L.A. et al. Cross-cohort gut microbiome associations with immune checkpoint inhibitor response in advanced melanoma. *Nat Med*. 2022;28(3):535-544
8. Olekhnovich E.I., Ivanov A.B., Babkina A.A. et al. Consistent Stool Metagenomic Biomarkers Associated with the Response To Melanoma Immunotherapy. *mSystems*. 2023;8(2):e0102322
9. Spencer C.N., McQuade J.L., Gopalakrishnan V. et al. Dietary fiber and probiotics influence the gut microbiome and melanoma immunotherapy response. *Science*. 2021;374(6575):1632-1640
10. Bowers R.M., Kyrpides N.C., Stepanauskas R. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*. 2017;35(8):725-731
11. Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353-D361
12. Caspi R., Billington R., Keseler I.M. et al. The MetaCyc database of metabolic pathways and enzymes – a 2019 update. *Nucleic Acids Res*. 2020;48(D1):D445-D453

doi 10.18699/bgrs2024-8.1-04

Кишечная микробиота при иммунотерапии меланомы: бактериофаги как прогностические биомаркеры и терапевтические агенты

Захаревич Н.*, Олехнович Е., Климина К.

Федеральный научно-клинический центр физико-химической медицины имени академика Ю.М. Лопухина ФМБА, Москва, Россия

** zakharevich@yandex.ru*

Ключевые слова: кишечная микробиота; меланома; вироm; бактериофаги; биомаркеры; иммунотерапия; метагеном

Мотивация и цель: Роль бактериофагов по влиянию на структуру и функции кишечной микробиоты (КМ) недооценена. В связи с быстрой эволюцией устойчивости бактерий к антибиотикам, фаготерапия является перспективным и актуальным подходом. Многие заболевания, в том числе и злокачественные опухоли, возникают и/или сопровождаются конкретными изменениями в КМ. Так, например, в последнее время КМ признана новым значимым игроком в патогенезе и лечении злокачественной меланомы. В ряде исследований было показано, что состав КМ значительно отличается у пациентов с меланомой по сравнению с контрольной группой, а также что КМ может влиять на противоопухолевый иммунитет у пациентов с меланомой и на эффективность иммунотерапии. Несмотря на это, исследования взаимодействия различных компонентов КМ и злокачественной меланомы все еще находятся в зачаточном состоянии – в связи с чем их изучение является актуальной задачей. В частности, необходимо с большей точностью выявлять различные микробные сигнатуры (в том числе патогенных бактерий), предрасполагающие к развитию побочных эффектов при заболевании и/или снижению эффективности иммунотерапии. Успешно снизить патогенную нагрузку и улучшить течение заболевания возможно, например, при специфическом нацеливании бактериофагов на конкретные бактериальные виды или даже штаммы. Помимо этого, известно, что фаги, наравне с бактериями, способны усиливать ответ на иммунотерапию индуцируя Т-клетки, перекрестно-реагирующие с раковыми антигенами. Исследования КМ, а именно кишечного вирома пациентов с меланомой и результаты, полученные в ходе данного исследования, способны обогатить терапевтический потенциал модулирования КМ и стимуляции системных противоопухолевых иммунных реакций, а также предложить новые неинвазивные биомаркеры заболевания – бактериофаги.

Методы и алгоритмы: С помощью инструмента fastq-dump (NCBI SRA Toolkit) сырые метагеномные данные были загружены из базы данных NCBI. Оценка качества полученных данных была выполнена с помощью программы FastQC. Тримминг данных был проведён при помощи инструмента Trimmomatic, а фильтрация последовательностей человеческой ДНК с использованием инструмента bbmap (версия генома человека GRCh37). Для реконструкции фаговых геномов из метагеномных последовательностей была проведена сборка сырых метагеномных данных (прочтений) в контиги параллельно с помощью двух

инструментов metaviralSPAdes и MEGAHIT. Для формирования качественного неизбыточного набора вирусных контигов (геномов) было выполнено предсказание последовательностей фагов и профагов в собранных метагеномных данных при помощи комбинации восьми алгоритмов: viralVerify, VirSorter2, DeepVirFinder, Seeker, VIBRANT, geNomad, DBSCAN-SWA и Phigaro – с последующим сравнением полученных результатов и отбором надежных предсказаний. Чтобы гарантировать качество и надежность идентифицированных вирусных контигов (геномов), был использован инструмент CheckV, позволяющий проверить качество и полноту предсказанных вирусных геномов. Последовательности с низким уровнем качества по результатам оценки инструмента CheckV были исключены из дальнейшего анализа. Затем с помощью инструмента dRep отобранные фаговые геномы были кластеризованы методом дерепликации. Полученный качественный неизбыточный каталог вирусных контигов был проанализирован с помощью библиотеки Python Phage BOX – являющейся набором инструментов для анализа фагов, включающим в себя: таксономическую классификацию вирусных контигов – инструмент PhaGCN, предсказание хозяев фагов – инструмент CHERRY, и определение типа взаимодействия бактериофага с бактериальной клеткой (вирулентный/ умеренный) – инструмент PhaTYP. Для выявления патогенных бактерий в собранных метагеномных данных была применена следующая стратегия анализа: в качестве референса была использована база данных патогенных организмов FDA-ARGOS, картирование собранных метагеномов осуществляли при помощи алгоритма HISAT2; при помощи пакета программ SAMtools, результаты картирования были конвертированы из SAM формата в BAM формат. Расчет относительных представленностей геномов патогенов производили при помощи программы CoverM. Обработка полученных результатов и расчет средней представленности бактерий проводили при помощи скриптов, написанных на языке R с использованием стандартных библиотек.

Результаты: В ходе проведённого исследования был сформирован каталог кишечных метагеномов от пациентов с меланомой с различным ответом на иммунотерапию. Этот каталог включает в себя 83 кишечных метагенома (образца) из двух опубликованных исследований [1, 2]. Анализируемые в проекте данные доступны в базе данных NCBI по следующим номерами доступа: PRJNA397906 и PRJNA399742. По результатам идентификации последовательностей бактериофагов в собранном каталоге кишечных метагеномов при помощи восьми алгоритмов для 83 образцов был предсказан 204751 потенциальный вирусный геном. После проверки качества и полноты предсказанных геномов осталось 182423 последовательности; а после этапа дерепликации было получено 46380 вирусных контигов – из которых и был сформирован окончательный неизбыточный качественный набор фаговых геномов. После таксономической аннотации полученного набора неизбыточных вирусных контигов получилось, что все последовательности принадлежат к классу *Caudoviricetes* и распределены по 19 различным семействам: *Peduoviridae* (10.38 %), *Straboviridae* (7.25 %), *Ackermannviridae* (3.71 %), *Salasmaviridae* (3.29 %), *Casjensviridae* (2.85 %), *Chaseviridae* (2.80 %), *Mesyanzhinovviridae* (2.11 %), *Herelleviridae* (1.84 %), *Drexlerviridae* (1.33 %), *Zierdtviridae* (0.94 %), *Kyanoviridae* (0.54 %), *Guelinviridae* (0.35 %), *Autographiviridae* (0.25 %), *Schitoviridae* (0.23 %), *Demerecviridae* (0.19 %), *Orlajensenviridae* (0.13 %), *Vilmaviridae* (0.11 %), *Rountreeviridae* (0.10 %),

Zobellviridae (0.01 %) и “no_family_avaliable” (1.88 %). Также для всех предсказанных вирусных контигов был определён тип взаимодействия с бактериальной клеткой – в результате для данных из исследования Frankel et al. [1] получилось, что вирулентные фаги составляют 32.95 %, а умеренные фаги 67.05 %; для данных из исследования Matson et al. [2] были получены сходные значения – 34.02 % для вирулентных фагов и 65.98 % для умеренных. Полученные результаты указывают на преобладание умеренных фагов в КМ, что соответствует опубликованным на сегодняшний день исследованиям. Как видно из описанного выше, разницы в процентном соотношении между вирулентными и умеренными фагами не наблюдалось, ни между наборами данных, ни между образцами метагеномов ответчиков (R) и неответчиков (NR). Проанализировав ассоциации между фагами и бактериальной составляющей КМ, были сделаны следующие выводы – во всех исследуемых образцах, независимо от набора данных, доминирующими хозяевами были десять бактериальных родов: *Clostridium*, *Faecalibacterium*, *Mycoplasma*, *Bifidobacterium*, *Ruminococcus*, *Bacteroides*, *Blautia*, *Streptomyces*, *Eubacterium* и семейство *Lachnospiraceae*. Что касается семейств фагов, то увеличение численности наблюдалось для четырёх, а именно: для набора образцов из исследования Frankel et al. [1] это были семейства *Drexlerviridae* (для ответчиков (R)) и *Salasmaviridae* (для неответчиков (NR)); для набора образцов из исследования Matson et al. [2] это были семейства *Chaseviridae* и *Mesyanzhinovviridae* (оба для неответчиков (NR)). Для всех фаговых семейств с увеличением численности, также наблюдался и более широкий спектр бактериальных хозяев. По предварительным результатам поиска бактериофагов – кандидатов в прогностические биомаркеры ответа на иммунотерапию было выделено два фаговых семейства *Drexlerviridae* и *Chaseviridae* в качестве наиболее перспективных кандидатов в прогностические биомаркеры. Однако для окончательных выводов необходимо провести более глубокий анализ большего числа данных. Что касается бактериальной составляющей КМ пациентов с меланомой, в 'Топ 5' патогенных видов с наибольшей средней представленностью среди исследованных образцов вошли такие бактерии как *Enterocloster bolteae*, *Clostridium scindens*, *Bacteroides fragilis*, *Raoultella ornithinolytica*, *Enterocloster clostridioformis*.

Выводы: На сегодняшний день нет работ, в которых бы анализировался виром ответивших и не ответивших на иммунотерапию онкологических больных. Наше исследование по возможности применения бактериофагов КМ в качестве прогностических биомаркеров исхода иммунотерапии ещё не закончено и мы планируем проанализировать больший объём данных и более глубоко, в том числе использовать в дальнейшем для поиска воспроизводимых таксономических прогностических биомаркеров метод дифференциального ранжирования с применением уже зарекомендовавшего себя в анализе сложных метаданных инструмента Songbird – этот метод позволяет оценивать относительную дифференциальную численность, независимо от общей микробной нагрузки. Таким образом, текущее исследование имеет шанс существенно расширить представления о вовлеченности КМ, в частности вирусной её составляющей, в иммунотерапию, а также предложить новые подходы к усовершенствованию существующих методов борьбы с злокачественной меланомой.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда (№ 23-75-10125, https://rscf.ru/project/23-75-10125/).

Gut microbiota in melanoma immunotherapy: bacteriophages as prognostic biomarkers and therapeutic agents

Zakharevich N.*, Olekhnovich E., Klimina K.
Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia
** zakharevich@yandex.ru*

Key words: gut microbiota; melanoma; virome; bacteriophages; biomarkers; immunotherapy; metagenome

Motivation and Aim: The role of bacteriophages in influencing the structure and function of the gut microbiota (GM) is underestimated. Due to the rapid evolution of bacterial resistance to antibiotics, phage therapy is a promising and relevant approach. Many diseases, including malignant tumors, occur and/or are accompanied by specific changes in GM. For example, recently GM has been recognized as a new significant player in the pathogenesis and treatment of malignant melanoma. In a number of studies, it has been shown that the composition of GM differs significantly in patients with melanoma compared to the control group, and also that GM can affect antitumor immunity in patients with melanoma and the effectiveness of immunotherapy. Despite this, research of the interaction of various components of GM and malignant melanoma are still in their infancy – and therefore their study is an urgent task. In particular, it is necessary to identify with greater accuracy various microbial signatures (including pathogenic bacteria) predisposing to the development of side effects in the disease and/or a decrease in the effectiveness of immunotherapy. It is possible to successfully reduce the pathogenic load and improve the course of the disease, for example, by specifically targeting bacteriophages to specific bacterial species or even strains. In addition, it is known that phages, along with bacteria, are able to enhance the response to immunotherapy by inducing T cells that cross-react with cancer antigens. Studies of GM, namely intestinal virome in patients with melanoma and the results obtained during this study, can enrich the therapeutic potential of modulating GM and stimulating systemic antitumor immune responses, as well as offer new non–invasive biomarkers of the disease – bacteriophages.
Methods and Algorithms: Using the fastq-dump tool (NCBI SRA Toolkit), raw metagenomic data was downloaded from the NCBI database. The evaluation of the quality of the obtained data was performed using the FastQC program. Data trimming was performed using the Trimomatic tool, and filtering of human DNA sequences using the bbmap tool (GRCh37 version of the human genome). To reconstruct phage genomes from metagenomic sequences, raw metagenomic data were assembled into contigs in parallel using two tools metaviralSPAdes and MEGAHIT. To form a qualitative, non-redundant set of viral contigs (genomes), the prediction of phages and prophages sequences in the collected metagenomic data was performed using a combination of eight algorithms: viralVerify, VirSorter2, DeepVirFinder, Seeker, VIBRANT, geNomad, DBSCAN–SWA and Phigaro, followed by a comparison of the results obtained and the selection of reliable predictions. To ensure the quality and reliability of the identified viral contigs (genomes), the CheckV tool was used to check the quality and completeness of the predicted viral genomes. Sequences with a low level of quality according to the results of the CheckV tool evaluation were excluded from further

analysis. Then, using the dRep tool, the selected phage genomes were clustered using the dereplication method. The resulting high–quality, non-redundant catalog of viral contigs was analyzed using the Python Phage BOX library, which is a set of tools for phage analysis, including: The PhaGCN tool provides taxonomic classification of viral contigs, the CHERRY tool predicts phage hosts, and the PhaTYP tool determines the type of interaction between a bacteriophage and a bacterial cell (virulent/temperate). To identify pathogenic bacteria in the collected metagenomic data, the following analysis strategy was applied: the FDA-ARGOS database of pathogenic organisms was used as a reference, the mapping of the collected metagenomes was carried out using the HISAT2 algorithm, using the SAMtools software package, the mapping results were converted from SAM format to BAM format. The relative representation of pathogen genomes was calculated using the CoverM program. The processing of the obtained results and the calculation of the average bacterial representation were carried out using scripts written in the R language using standard libraries.

Results: During the study, a catalog of gut metagenomes from melanoma patients with different responses to immunotherapy was formed. This catalog includes 83 gut metagenomes (samples) from two published studies [1, 2]. The data analyzed in the project is available in the NCBI database using the following access numbers: PRJNA397906 and PRJNA399742. Based on the results of the identification of bacteriophage sequences in the collected catalog of gut metagenomes, 204,751 potential viral genomes were predicted using eight algorithms for 83 samples. After checking the quality and completeness of the predicted genomes, 182,423 sequences remained; and after the dereplication stage, 46,380 viral contigs were obtained – from which the final non-redundant qualitative set of phage genomes was formed. After taxonomic annotation of the obtained set of non-redundant viral contigs, it turned out that all sequences belong to the class *Caudoviricetes* and are distributed into 19 different families: *Peduoviridae* (10.38 %), *Straboviridae* (7.25 %), *Ackermannviridae* (3.71 %), *Salasmaviridae* (3.29 %), *Casjensviridae* (2.85 %), *Chaseviridae* (2.80 %), *Mesyanzhinovviridae* (2.11 %), *Herelleviridae* (1.84 %), *Drexlerviridae* (1.33 %), *Zierdtviridae* (0.94 %), *Kyanoviridae* (0.54 %), *Guelinviridae* (0.35 %), *Autographiviridae* (0.25 %), *Schitoviridae* (0.23 %), *Demerecviridae* (0.19 %), *Orlajensenviridae* (0.13 %), *Vilmaviridae* (0.11 %), *Rountreeviridae* (0.10 %), *Zobellviridae* (0.01 %) and “no_family_avaliable” (1.88 %). Also, for all predicted viral contigs, the type of interaction with a bacterial cell was determined – as a result, for data from the study by Frankel et al. [1], it turned out that virulent phages account for 32.95 %, and temperate phages 67.05 %; for data from the study by Matson et al. [2], similar values were obtained – 34.02 % for virulent phages and 65.98 % for temperate ones. The results indicate a predominance of temperate phages in GM, which is consistent with studies published to date. As can be seen from the above, there was no difference in percentage between virulent and temperate phages, neither between datasets nor between responder (R) and non-responders (NR) metagenome samples. Having analyzed the associations between phages and the bacterial component of GM, the following conclusions were made – in all studied samples, regardless of the data set, the dominant hosts were ten bacterial genera: *Clostridium*, *Faecalibacterium*, *Mycoplasma*, *Bifidobacterium*, *Ruminococcus*, *Bacteroides*, *Blautia*, *Streptomyces*, *Eubacterium* and family *Lachnospiraceae*. Regarding phage families, an increase in abundance was observed for four, namely: for the set of samples from the study of Frankel et al. [1], these were the families *Drexlerviridae* (for responders (R)) and

Salasmaviridae (for non-responders (NR)); for the set of samples from the study by Matson et al. [2], these were the families *Chaseviridae* and *Mesyanzhinovviridae* (both for non-responders (NR)). For all phage families, as abundance increased, a wider range of bacterial hosts was also observed. Based on preliminary results of the search for bacteriophages – candidates for predictive biomarkers of response to immunotherapy, two phage families, *Drexlerviridae* and *Chaseviridae*, were identified as the most promising candidates for predictive biomarkers. However, for final conclusions it is necessary to conduct a more in-depth analysis of a larger amount of data. As for the bacterial component of GM from patients with melanoma, the "Top 5" pathogenic species with the highest average representation among the studied samples included bacteria such as *Enterocloster bolteae*, *Clostridium scindens*, *Bacteroides fragilis*, *Raoultella ornithinolytica*, *Enterocloster clostridioformis*.

Conclusion: To date, there are no studies that would analyze the virome of cancer patients who responded and did not respond to immunotherapy. Our study on the possibility of using GM bacteriophages as prognostic biomarkers of the outcome of immunotherapy has not yet been completed and we plan to analyze a larger volume of data and in more depth, in particular, in the future, to use the differential ranking method for searching for reproducible taxonomic prognostic biomarkers using the Songbird tool, which has already proven itself in the analysis of complex metadata – this method allows you to estimate the relative differential abundance, regardless of the total microbial load. Thus, the current study has the chance to significantly expand the understanding of the involvement of GM, in particular its viral component, in immunotherapy, as well as to propose new approaches to improving existing methods of combating malignant melanoma.

Funding: The research was funded by the Russian Science Foundation (No. 23-75-10125, https://rscf.ru/project/23-75-10125/).

Список литературы/References

1. Frankel A.E., Coughlin L.A., Kim J., Froehlich T.W., Xie Y., Frenkel E.P., Koh A.Y. Metagenomic shotgun sequencing and unbiased metabolomic profiling identify specific human gut microbiota and metabolites associated with immune checkpoint therapy efficacy in melanoma patients. *Neoplasia*. 2017;19(10):848-855. doi 10.1016/j.neo.2017.08.004
2. Matson V., Fessler J., Bao R., Chongsuwat T., Zha Y., Alegre M.-L., Luke J.J., Gajewski T.F. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science*. 2018;359(6371):104-108. doi 10.1126/science.aao3290

doi 10.18699/bgrs2024-8.1-05

Поиск универсальных метагеномных маркеров микробиоты кишечника, ассоциированных с ответом на иммунотерапию различных видов рака

Канаева В.[1, 2]*, Климина К.[1], Олехнович Е.[1]

[1] *Федеральный научно-клинический центр физико-химической медицины имени академика Ю.М. Лопухина Федерального медико-биологического агентства, Москва, Россия*

[2] *Московский физико-технический институт (национальный исследовательский университет), Москва, Россия*

* *vera.a.kanaeva@gmail.com*

Ключевые слова: кишечная микробиота; метагеномные маркеры; предсказательная модель; иммунотерапия; различные виды рака

Мотивация и цель: Известно, что структура и состав микробиоты кишечника человека влияют на состояние его иммунной системы [1]. Более того, кишечные микробы могут воздействовать на противоопухолевый иммунитет и повышать эффективность иммунотерапии, основанной на ингибировании иммунных контрольных точек. Настоящее исследование направлено на определение универсальных метагеномных маркеров микробиоты кишечника, которые коррелируют с успешностью иммунотерапии при различных видах рака.

Методы и алгоритмы: Для данной работы были загружены метагеномы из 11 опубликованных независимых исследований из открытой базы данных SRA-NCBI. Всего было собрано 814 образцов кишечной микробиоты пациентов, страдающих различными онкологическими заболеваниями, включая меланому, рак желудочно-кишечного тракта (толстой и прямой кишки, поджелудочной железы), рак легких, молочных желез и яичников. 462 образца от пациентов, которые ответили на терапию и 352 от не ответивших.

Для всех образцов проведен контроль качества прочтений (fastQC) и фильтрация человеческой ДНК (kraken2). Первый этап нашей работы заключался в создании неизбыточного каталога геномов, собранных из метагеномов (metagenome-assembled genome – MAG). Данный процесс включал сборку метагеномных контигов (megahit), биннинг (metabat2, maxbin2), дерепликацию (dRep) собранных метагеномных бинов и их таксономическую (GTDB-TK) аннотацию. Вторым этапом было выявление метагеномных маркеров. Для этого был получен профиль представленности MAG`ов в образцах (InStrain) и проведено ранжирование по степени влияния на исход иммунотерапии (SongBird). Гены, обнаруженные в маркерных геномах, были проаннотированы базой данных KEGG.

Результаты: В результате создания каталога MAG’ов было собрано 3855 операционных геномных единиц (ОГЕ) из которых было выявлено 424 метагеномных маркера: 166 связаны с позитивным исходом лечения, а 258 – с негативным.

Среди маркеров, ассоциированных с отсутствием ответа на иммунотерапию были выявлены патогенные микроорганизмы *Raoultella ornithinolytica*, *Haemophilus_D* и *Klebsiella michiganensis*. *Hungatella effluvii*, чьё присутствие в кишечнике

человека связывают с тяжелым течением саркопении [2], также обнаружена в числе отрицательных маркеров. Среди маркеров, ассоциированных с успешным исходом лечения, наиболее часто встречаются *Bifidobacterium adolescentis*, *Bacteroides uniformis*, *Alistipes putredinis*. Эти виды способствуют увеличению числа симбиотических бактерий в кишечнике, стимулируют сохранение кишечного барьера, обладают иммуномодулирующим действием и ассоциированы с благоприятными исходами лечения различных заболеваний [3–5].

В результате анализа функциональных путей, было установлено, что у пациентов, положительно реагирующих на иммунотерапию, по сравнению с теми, кто не показал ответа, наблюдалось статистически значимое увеличение метаболизма крахмала и сахарозы. Эти вещества активно ферментируются *B. adolescentis*, которая также участвует в ферментации других гликанов [3]. В то же время успешный исход иммунотерапии связан с уменьшением активности пути образования биопленки *Escherichia coli*, которая часто выступает в роли патогенного фактора.

Для построения предсказательной модели использовались данные о логарифмическом отношении относительных представленностей выявленных маркерных ОГЕ в каждом исследуемом образце. Применение модели логистической регрессии позволило спрогнозировать ответ на иммунотерапию на основании лог-отношений для различных видов рака. Разработанная модель продемонстрировала улучшенное качество предсказания по сравнению с традиционными методами таксономической аннотации и относительных представленностей, принятыми в научном сообществе по изучению микробиома.

Выводы: Результаты данной работы создают основу для формирования гипотезы о влиянии микробиоты кишечника на эффективность иммунотерапии рака. Введение нового метода анализа метагеномных данных обеспечивает более детальное описание состава кишечной микробиоты и повышает точность прогнозирования исходов лечения. Это исследование способствует переходу к персонализированной медицине и разработке методов предсказания реакции на иммунотерапию у пациентов с различными видами рака.

Финансирование: Исследование поддержано грантом РНФ (№ 22-75-10029, https://rscf.ru/project/22-75-10029/).

Search for universal metagenomic markers of the gut microbiota associated with the response to immunotherapy of various types of cancer

Kanaeva V.[1, 2]*, Klimina K.[1], Olekhnovich E.[1]

[1] *Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, Moscow, Russia*

* *vera.a.kanaeva@gmail.com*

Key words: intestinal microbiota; metagenomic markers; predictive model; immunotherapy; various types of cancer

Motivation and Aim: It is known that the structure and composition of the human gut microbiota affect the state of the human immune system [1]. Moreover, intestinal microbes can affect antitumor immunity and increase the effectiveness of immunotherapy based on inhibition of immune checkpoints. The present study aims to identify universal metagenomic markers of the gut microbiota that correlate with the success of immunotherapy in various types of cancer.
Methods and Algorithms: For this work, metagenomes from 11 published independent studies from the open SRA-NCBI database were downloaded. A total of 814 samples of the intestinal microbiota of patients suffering from various oncological diseases, including melanoma, cancer of the gastrointestinal tract (colon, rectum, pancreas), lung, breast and ovarian cancers, were collected. 462 samples from patients who responded to therapy and 352 from non-responders.
Reads quality control (fastQC) and human DNA filtration (kraken2) were performed for all samples. The first stage of our work was to create an endless catalog of genomes assembled from metagenomes (metagenome-assembled genome – MAG). This process included the assembly of metagenomic contigs (megahit), binning (metabat2, maxbin2), dereplication (dRep) of the collected metagenomic bins and their taxonomic (GTDB-TK) annotation. The second stage was the identification of metagenomic markers. To do this, a profile of MAGS in the samples (InStrain) was obtained and a ranking was carried out according to the degree of influence on the outcome of immunotherapy (SongBird). The genes found in the marker genomes were annotated by the KEGG database.
Results: The MAG's catalog resulted in the collection of 3,855 operational genomic units (OGEs) of which 424 metagenomic markers were identified: 166 were associated with positive treatment outcome and 258 with negative treatment outcome.
The pathogens *Raoultella ornithinolytica*, *Haemophilus_D* and *Klebsiella michiganensis* were identified among the markers associated with lack of response to immunotherapy. *Hungatella effluvii*, whose presence in the human gut has been associated with the severe course of sarcopenia [2], was also found among the negative markers. Among the markers associated with successful treatment outcome, *Bifidobacterium adolescentis*, *Bacteroides uniformis*, and *Alistipes putredinis* are the most common. These species promote an increase in the number of symbiotic bacteria in the gut, stimulate the maintenance of the intestinal barrier, have immunomodulatory effects, and are associated with favorable treatment outcomes for various diseases [3–5].
Through functional pathway analysis, it was found that patients who responded positively to immunotherapy, compared to those who did not show a response, showed a statistically significant increase in starch and sucrose metabolism. These substances are actively fermented by B. adolescentis, which is also involved in the fermentation of other glycans [3]. At the same time, the successful outcome of immunotherapy is associated with a decrease in the activity of the Escherichia coli biofilm formation pathway, which often acts as a pathogenic factor.
Data on the log-ratio of the relative representations of the identified marker OGEs in each study sample were used to build a predictive model. Application of a logistic regression model allowed predicting the response to immunotherapy based on log-ratios for different cancer types. The developed model demonstrated improved prediction quality compared to traditional methods of taxonomic annotation and relative representations accepted in the microbiome research community.
Conclusion: The results of this work provide a basis for the formation of a hypothesis on the influence of the gut microbiota on the efficacy of cancer immunotherapy. The

introduction of a new method of analyzing metagenomic data provides a more detailed description of the composition of the gut microbiota and improves the accuracy of predicting treatment outcomes. This study contributes to the transition to personalized medicine and the development of methods to predict response to immunotherapy in patients with different types of cancer.

Funding: The study is supported by the RSF (No. 22-75-10029, https://rscf.ru/project/22-75-10029/).

Список литературы/References

1. Gomaa E.Z. Human gut microbiota/microbiome in health and diseases: a review. *Antonie Van Leeuwenhoek*. 2020;113:2019-2040. doi 10.1007/s10482-020-01474-7
2. Wang Y. et al. Population-based metagenomics analysis reveals altered gut microbiome in sarcopenia: data from the Xiangya Sarcopenia Study. *J Cachexia Sarcopenia Muscle*. 2022;13:2340-2351. doi 10.1002/jcsm.13037
3. Leser T., Baker A. Bifidobacterium adolescentis – a beneficial microbe. *Beneficial Microbes*. 2023;14(6):525-551. doi 10.1163/18762891-20230030
4. Yan Y. et al. Bacteroides uniformis-induced perturbations in colonic microbiota and bile acid levels inhibit TH17 differentiation and ameliorate colitis developments. *NPJ Biofilms Microbiomes*. 2023;9(1):56. doi 10.1038/s41522-023-00420-5
5. Parker B. et al. The genus *Alistipes*: gut bacteria with emerging implications to inflammation, cancer, and mental health. *Front Immunol*. 2020;11:906. doi 10.3389/fimmu.2020.00906

doi 10.18699/bgrs2024-8.1-06

Влияние лактобацилл и бифидобактерий на кишечную микробиоту мышей: сравнительный анализ методов секвенирования с использованием платформ Illumina и Oxford Nanopore Technologies

Климина К.[1]*, Зорук П.[1], Строкач А.[1], Веселовский В.[1], Бабенко В.[1], Колдман С.[1, 3], Колдман В.[1, 3], Одорская М.[2], Даниленко В.[2], Селезнева О.[1], Захаревич Н.[1], Ларин А.[1], Морозов М.[1], Олехнович Е.[1]

[1] *Федеральный научно-клинический центр Физико-химической медицины имени академика Ю.М. Лопухина Федерального медико-биологического агентства, Москва, Россия*

[2] *Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия*

[3] *ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России, Москва, Россия*

** ppp843@yandex.ru*

Ключевые слова: Микробиота кишечника; Oxford Nanopore Technologies; Illumina; бактериальное разнообразие; 16S рРНК; *Lacticaseibacillus rhamnosus*; *Bifidobacterium adolescentis*

Мотивация и цель: Микробиом кишечника является одним из крупнейших бактериальных сообществ в человеческом организме и влияет как на поддержание здоровья, так и на развитие различных заболеваний [1]. Известно, что комменсальные бактерии, такие как бифидобактерии и лактобациллы, оказывают благотворное влияние на состояние организма хозяина, модулируя иммунные реакции, укрепляя кишечный барьер и подавляя рост патогенных микроорганизмов [2]. Таким образом, изучение микробиоты кишечника является важной и перспективной областью исследований. Технологии секвенирования нового поколения произвели революцию в изучении микробиоты, так как с помощью них была получена детальная информация о микробных сообществах. В последние годы был достигнут большой прогресс в определении состава бактериальных сообществ кишечника благодаря развитию технологий секвенирования с использованием приборных комплексов различных типов таких как Illumina и Oxford Nanopore Technologies (ONT). Основными методами для идентификации микробных сообществ служат анализ последовательности гена 16S рРНК и полногеномное секвенирование (whole genome sequencing (WGS)). Цель данного исследования заключается в изучении воздействия штаммов *Lacticaseibacillus rhamnosus* K32 и *Bifidobacterium adolescentis* 150 на микробиоту кишечника мышей, а также сравнение методов секвенирования на платформах Illumina и ONT для определения наиболее эффективного подхода к анализу состава микробиоты кишечника.

Методы и алгоритмы: 36 самок мышей линии C57BL/6 были приобретены в филиале «Столбовая». Животных содержали в карантине в течение 14 дней, затем их случайным образом разделили на три группы: контрольная группа; мыши, получавшие лактобациллы; мыши, получавшие бифидобактерии. Образцы фекалий мышей собирали в стерильный эппендорф на 0, 5, 8, 12, 15, 19, 21, 25, 28 дни эксперимента. Образцы хранили при температуре –80 °C. Штамм

L. rhamnosus K32 (GenBank JNNV00000000) выращивали в анаэробных условиях (10 % CO2, Anaerobic System Mark II, HiMedia, Мумбаи, Индия). Штамм *B. adolescentis* 150 (GenBank LBHQ00000000) также выращивали в анаэробных условиях с добавлением 0.5 % цистеина. Штаммы культивировали при температуре +37 °C на среде MRS (HiMedia). ДНК выделяли с использованием набора PureLink™ Microbiome DNA Purification Kit (Thermo Fisher Scientific, США) и Wizard Genomic DNA Purification Kit (Promega, США). ДНК амплифицировали с использованием праймеров 27F, bif27F, 8F, 1492R с помощью набора Tersus Plus PCR в 25 мкл. Библиотеки для ONT готовили в соответствии с протоколом (Ligation sequencing amplicons). Секвенирование гена 16S rRNA проводили на платформе MinION на ячейке R9.4.1 (FLO-MIN106; ONT). Использовалось программное обеспечение MINKNOW версии 22.12.7 (ONT). Подготовка библиотеки на регион V3-V4 гена 16S рРНК и секвенирование проводились в соответствии с протоколом Illumina [3]. Для подготовки WGS библиотек использовали 100 нг ДНК и набор KAPA HyperPlus Kit (Roche, Швейцария). Секвенирование ДНК-библиотек проводили на платформе HiSeq 2500 (Illumina, США) в соответствии с рекомендациями производителя, используя следующие наборы реагентов: HiSeq Rapid PE Cluster Kit v2, HiSeq Rapid SBS Kit v2 (200 циклов) и HiSeq Rapid PE FlowCell v2 и 2 % PhiX. Длинные риды получали с помощью секвенирования на PromethION (ONT), Великобритания) согласно протоколу производителя, с реагентами NEB. Полученные риды обрабатывали с использованием Guppy v6.5.7. с использованием стандартных параметров. Fastp использовали для обрезки последовательностей низкого качества и фильтрации технических последовательностей [4]. Анализ данных проводили с использованием DADA2 pipeline [5] и базы данных SILVA [6]. Различия между сиквенсовыми технологиями определяли с помощью LefSe [7].

Результаты: В ходе эксперимента было исследовано влияние двух пробиотических штаммов, *L. rhamnosus* K32 и *B. adolescentis* 150, на кишечную микробиоту мышей линии C57BL/6. Изменения бактериального состава оценивали путем секвенирования гена 16S рРНК и WGS на двух платформах – Illumina и ONT. Для секвенирования гена 16S рРНК на ONT использовали 5 известных комбинаций праймеров, и проводили сравнительный анализ эффективности определения бактериального разнообразия с помощью данных комбинаций. Все 5 исследуемых комбинаций праймеров на ген 16S рРНК позволили выявить сопоставимое альфа-разнообразие кишечной микробиоты мышей. Мы провели попарное сравнение данных, полученных с разных платформ для секвенирования и с разной исходной ДНК (обычная и высокомолекулярная). Результаты исследования демонстрируют высокую степень корреляции между данными, полученными с одной платформы и с разным типом исходной ДНК. Однако данные WGS, полученные с платформ Illumina и ONT, имеют высокий уровень сходства. В целом, использование комбинированных данных с двух платформ секвенирования позволяет более полно и точно оценивать бактериальное разнообразие в анализируемых образцах. Введение бактериальных штаммов *L. rhamnosus* K32 и *B. adolescentis* 150 значительно влияет на состав микробиоты у мышей. Статистический анализ с использованием индекса Шеннона выявил существенные различия между контрольной группой и группами, которым вводили лиофилизированные культуры этих штаммов. Наблюдаемое увеличение альфа-разнообразия в экспериментальных группах

может свидетельствовать о потенциально благоприятном влиянии лечения на микробиоту кишечника. В ходе комплексного исследования микробиоты мышей, которым вводили штаммы *L. rhamnosus* K32 и *B. adolescentis* 150, было показано, что оба штамма оказывают положительное влияние на микробиоту кишечника. Примечательно, что введение этих штаммов привело к заметному увеличению популяции рода *Anaerotignum*, включая такие виды, как *Anaerotignum lactatifermentans, Anaerotignum aminivorans* и *Anaerotignum faecicola,* которые известны выработкой полезных метаболитов, оказывающих благотворное влияние на пищеварение: пропионата, ацетата и бутирата [8, 9]. Также в кишечнике мышей, получавших лактобациллы и бифидобактерии, наблюдалось значительное увеличение *Bacteroides acidifaciens*, которая может способствовать усилению барьерной функции кишечника и снижению интенсивности воспалительных процессов [10, 11].

Выводы: Наблюдаемые изменения в сочетании с общим физическим самочувствием мышей свидетельствуют о благотворном влиянии введенных бактериальных культур *L. rhamnosus* K32 и *B. adolescentis* 150 на здоровье. Данное исследование подчеркивает важность выбора подходящих стратегий секвенирования для тщательного изучения микробиома. Использование гибридного подхода к сборке увеличивает количество идентифицированных контигов.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 23-75-10125 https://rscf.ru/project/23-75-10125/.

Impact of lactobacilli and bifidobacteria on mouse gut microbiota: comparative analysis of sequencing methods using Illumina and Oxford Nanopore Technologies platforms

Klimina K.[1]*, Zoruk P.[1], Strokach A.[1], Veselovsky V.[1], Babenko V.[1], Koldman S.[1, 3], Koldman V.[1, 3], Odorskaya M.[2], Danilenko V.[2], Selezneva O.[1], Zakharievich N.[1], Larin A.[1], Morozov M.[1], Olekhnovich E.[1]

[1] *Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine, Moscow, Russia*

[2] *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia*

[3] *Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russia*

* *ppp843@yandex.ru*

Key words: Gut microbiota; Oxford Nanopore Technologies; Illumina sequencing; microbial diversity; 16S rRNA gene sequencing; *Lacticaseibacillus rhamnosus*; *Bifidobacterium adolescentis*

Motivation and Aim: The gut microbiome is one of the largest bacterial communities in the human body and influences both health maintenance and the development of various diseases [1]. It is known that commensal bacteria, such as bifidobacteria and lactobacilli, have a beneficial effect on the host's health by modulating immune responses, strengthening the intestinal barrier, and suppressing the growth of pathogenic microorganisms [2]. Therefore, the study of the gut microbiota is an important and promising field of research. Next-generation sequencing technologies have revolutionized the study of microbiota, providing detailed information about microbial communities. In recent years, significant progress has been made in determining the

composition of gut bacterial communities thanks to the development of sequencing technologies using various types of instrument platforms such as Illumina and Oxford Nanopore Technologies (ONT). The main methods for identifying microbial communities are 16S rRNA gene sequencing and whole-genome sequencing (WGS). The goal of this study is to investigate the impact of strains *Lacticaseibacillus rhamnosus* K32 and *Bifidobacterium adolescentis* 150 on the gut microbiota of mice, as well as to compare sequencing methods on the Illumina and ONT platforms to determine the most effective approach for analyzing the composition of the gut microbiota.

Methods and Algorithms: Thirty-six female C57BL/6 mice were acquired from the "Stolbovaya" branch. The animals were kept in quarantine for 14 days, after which they were randomly divided into three groups: a control group; mice receiving lactobacilli; and mice receiving bifidobacteria. Mouse fecal samples were collected into sterile Eppendorf tubes on days 0, 5, 8, 12, 15, 19, 21, 25, and 28 of the experiment and stored at –80 °C. The strain *L. rhamnosus* K32 (GenBank JNNV00000000) was cultured under anaerobic conditions (10 % CO2, Anaerobic System Mark II, HiMedia, Mumbai, India). The strain *B. adolescentis* 150 (GenBank LBHQ00000000) was also cultured under anaerobic conditions with the addition of 0.5 % cysteine. Both strains were cultivated at +37 °C in MRS medium (HiMedia). DNA was extracted using the PureLink™ Microbiome DNA Purification Kit (Thermo Fisher Scientific, USA) and the Wizard Genomic DNA Purification Kit (Promega, USA). DNA was amplified using primers 27F, bif27F, 8F, and 1492R with the Tersus Plus PCR kit in a 25 µl reaction volume. Libraries for ONT were prepared according to the ligation sequencing amplicons protocol. 16S rRNA gene sequencing was performed on the MinION platform using the R9.4.1 flow cell (FLO-MIN106; ONT) with MINKNOW software version 22.12.7 (ONT). The preparation of libraries targeting the V3-V4 region of the 16S rRNA gene and sequencing were conducted according to the Illumina protocol [3]. For WGS library preparation, 100 ng of DNA was used with the KAPA HyperPlus Kit (Roche, Switzerland). DNA library sequencing was performed on the HiSeq 2500 platform (Illumina, USA) following the manufacturer's recommendations using the HiSeq Rapid PE Cluster Kit v2, HiSeq Rapid SBS Kit v2 (200 cycles), and HiSeq Rapid PE Flow Cell v2, with a 2 % PhiX spike-in. Long reads were generated using PromethION sequencing (ONT, UK) according to the manufacturer's protocol with NEB reagents. The reads were processed using Guppy v6.5.7 with standard parameters. Fastp was used for trimming low-quality sequences and filtering technical sequences [4]. Data analysis was conducted using the DADA2 pipeline [5] and the SILVA database [6]. Differences between sequencing technologies were identified using LefSe [7].

Results: During the experiment, the impact of two probiotic strains, *L. rhamnosus* K32 and *B. adolescentis* 150, on the gut microbiota of C57BL/6 mice was investigated. Changes in the bacterial composition were assessed through 16S rRNA gene sequencing and whole-genome sequencing (WGS) on two platforms – Illumina and ONT. For 16S rRNA gene sequencing on ONT, five known primer combinations were used, and a comparative analysis of the effectiveness of these combinations in determining bacterial diversity was conducted. All five primer combinations for the 16S rRNA gene allowed for the detection of comparable alpha-diversity in the gut microbiota (GM) of the mice. Pairwise comparisons of data obtained from different sequencing platforms and from different types of DNA (standard and high molecular weight) were conducted. The results demonstrate a high degree of correlation between data obtained from the same platform with different types of starting DNA. However, WGS data obtained from Illumina

and ONT platforms show a high level of similarity. Overall, using combined data from two sequencing platforms allows for a more comprehensive and accurate assessment of bacterial diversity in the analyzed samples. The introduction of bacterial strains *L. rhamnosus* K32 and *B. adolescentis* 150 significantly influences the composition of the microbiota in mice. Statistical analysis using the Shannon index revealed significant differences between the control group and the groups that were administered lyophilized cultures of these strains. The observed increase in alpha-diversity in the experimental groups may indicate a potentially beneficial effect of treatment on gut microbiota. In the comprehensive study of the microbiota of mice administered strains *L. rhamnosus* K32 and *B. adolescentis* 150, it was shown that both strains have a positive impact on the GM. Notably, the introduction of these strains led to a significant increase in the population of the genus *Anaerotignum*, including species such as *Anaerotignum lactatifermentans, Anaerotignum aminivorans,* and *Anaerotignum faecicola*, which are known for producing beneficial metabolites that positively affect digestion: propionate, acetate, and butyrate [8, 9]. Additionally, in the intestines of mice that received lactobacilli and bifidobacteria, there was a significant increase in *Bacteroides acidifaciens*, which may enhance the barrier function of the intestine and reduce the intensity of inflammatory processes [10, 11].

Conclusion: The observed changes, combined with the overall physical well-being of the mice, suggest a beneficial impact of the introduced bacterial cultures *L. rhamnosus* K32 and *B. adolescentis* 150 on health. This study highlights the importance of choosing appropriate sequencing strategies for thorough microbiome research. The use of a hybrid assembly approach increases the number of identified contigs.

Funding: Financial support for this study was provided by the Russian Science Foundation under the grant number № 23-75-10125 https://rscf.ru/project/23-75-10125/.

Список литературы/References

1. Blaut M., Clavel T. Metabolic diversity of the intestinal microbiota: implications for health and Disease. *J Nutr*. 2007;137(3):751S-755S. doi 10.1093/jn/137.3.751S
2. O'Toole P.W., Marchesi J.R., Hill C. Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. *Nat Microbiol*. 2017;2(5):16221
3. 16S metagenomic libraries preparation. Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System. Part # 15044223. Rev. B. 2013. [https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf]
4. Chen Sh. et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884-i890. doi 10.1093/bioinformatics/bty560
5. Callahan B., McMurdie P., Rosen M. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13:581-583. doi 10.1038/nmeth.3869
6. Quast C., Pruesse E. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41:D590-D596. doi 10.1093/nar/gks1219
7. Segata N., Izard J., Waldron L., Gevers D., Miropolsky L., Garrett W.S., Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12(6):R60. doi 10.1186/gb-2011-12-6-r60
8. Hosseini E., Grootaert C., Verstraete W., Van de Wiele T. Propionate as a health-promoting microbial metabolite in the human gut. *Nutr Rev*. 2011;69(5):245-258. doi 10.1111/j.1753-4887.2011.00388.x
9. Ueki A., Goto K., Ohtaki Y. et al. Description of Anaerotignum aminivorans gen. nov., sp. nov., a strictly anaerobic, amino-acid-decomposing bacterium isolated from a methanogenic reactor, and reclassification of Clostridium propionicum, Clostridium neopropionicum and Clostridium lactatifermentans as species of the genus Anaerotignum. *Int J Syst Evol Microbiol*. 2017;67(10):4146-4153. doi 10.1099/ijsem.0.002268
10. Yang J.Y., Lee Y.S., Kim Y. et al. Gut commensal Bacteroides acidifaciens prevents obesity and improves insulin sensitivity in mice. *Mucosal Immunol*. 2017;10(1):104-116. doi 10.1038/mi.2016.42
11. Zheng C., Zhong Y., Xie J. et al. Bacteroides acidifaciens and its derived extracellular vesicles improve DSS-induced colitis. *Front Microbiol.* 2023;14:1304232. doi 10.3389/fmicb.2023.1304232

doi 10.18699/bgrs2024-8.1-07

Разработка алгоритмов и подходов для изучения перспективных штаммов-фармабиотиков с использованием омиксных технологий

Марсова М.[1]*, Резникова Д.[1, 2], Летвинова В.[1], Галанова О.[1, 2]

[1] *Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия*

[2] *Московский физико-технический институт (национальный исследовательский университет) (МФТИ), Москва, Россия*

* *masha_marsova@mail.ru*

Ключевые слова: омиксные технологии; фармабиотики; протеомный анализ; антиоксиданты; микробиом

Мотивация и цель: Различия между микробиомами клинически здоровых людей и людей, с проблемами здоровья, могут служить как диагностическим признаком, так и целью для последующей терапии. В свете неутешительной статистики роста случаев неинфекционных хронических заболеваний, возможность коррекции микробиома приобретает новые смыслы. При этом перспективы практического применения традиционных пре- и пробиотиков остаются размытыми. На смену им должно прийти новое поколение стандартизированных препаратов, нацеленных на коррекцию конкретных нозологий.

Цель – разработка принципиально новых подходов и алгоритмов исследования перспективных штаммов-фармабиотиков, позволяющих выявить молекулярные факторы и механизмы, обеспечивающие их клиническую эффективность.

Методы и алгоритмы: Для исследования, в качестве перспективного фармабиотика, был выбран штамм *Levilactobacillus brevis* 47f, проявивший наибольшую среди родственных штаммов адаптогенную и антиоксидантную активность в моделях *in vitro* (при помощи биолюминесцентных тестов) и *in vivo* (на моделях индуцированного мукозита у мышей, стресса, вызванного у мышей и низкодозового воздействия ксенобиотиков у рыб *Danio rerio*) [1–4].

Для выявления действующего начала *L. brevis* 47f применялись как традиционные биологические модели и сравнительный геномный анализ, так и омиксные подходы: протеомные, транскриптомные и метаболомные методы.

В исследовании были использованы результаты сравнительного геномного анализа *L. brevis* 47f (WGS PRJNA280953 в базе GenBank) и родственных штаммов из базы NCBI, а также каталог генов, способных влиять на проявление антиоксидантной активности [5, https://github.com/Alexey-Kovtun/Catalog].

Тотальную РНК для транскриптомного анализа выделяли на автоматизированной станции King Fisher с помощью MagMax™ mirVana™. Готовые библиотеки секвенировали на приборе Illumina HiSeq 2500. Для составления карты считываний и оценки количества транскриптов использовалось программное обеспечение Kallisto версии 0.46.0. Дифференциальный анализ экспрессии проводили с помощью edgeR пакет v3.26.8 интегрирован в веб-инструмент Degust v4.1.1.

Хромато-масс-спектрометрический анализ внутриклеточных белков проводили при помощи обращенно-фазной хроматографии на приборе Ultimate 3000 Nano LC System (Thermo Fisher Scientific), соединенным с масс-спектрометром Q Exactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific) посредством наноэлектроспрейного источника (Thermo Fisher Scientific). Полученные данные анализировали при помощи программы Peaks studio 10.0 (Bioinformatics Solutions Inc.). Идентификацию белков проводили посредством корреляции тандемных масс-спектров с базой данных белковых последовательностей из базы Uniprot.
Летучие метаболиты культуральной жидкости анализировали на приборе ГХ/МС Shimadzu QP2010 Ultra с парофазным экстрактором Shimadzu HS-20, колонкой VF-WAX MS. Для обработки хроматограммы использовали ПО Shimadzu LabSolutions и библиотеки масс-спектров NIST14 с пакетом AMDIS, который обнаруживает соединения на хроматограмме и получает очищенные масс-спектры для каждого соединения.
В качестве дополнительных методов проводили гель-электрофорез внутриклеточных белков и белков культуральной жидкости с масс-спектрометрическим исследованием характерных белковых фракций.
Данные протеомного, транскриптомного и метаболомного анализа были проанализированы при помощи разработанных биоинформатических алгоритмов с использованием языка программирования Python и машинного обучения. Полученные результаты сопоставлялись с данными стандартных физико-химических методов.
Результаты: Сравнительный геномный анализ позволил выявить ряд не характерных для *Levilactobacillus brevis* генов, продукты которых способны влиять на проявляемые в биологических моделях свойства штамма. Например, глутаредоксин-подобный белок NrdH, глутатион редуктаза/НАД(Ф)/ФАД-зависимая оксидоредуктаза, Dyp-type пероксидаза и др. Транскрипты соответствующих генов, а также белки были обнаружены в результатах транскриптомного и протеомного анализа соответственно. Сопоставление результатов, полученных при помощи омиксных технологий и физико-химических методов анализа, позволяет сделать предположения о возможных механизмах и метаболитах, вовлеченных в осуществление наблюдаемой антиоксидантной активности. В частности, некоторые выявленные белки и метаболиты могут проявлять не только антиоксидантную, но и иммуномодулирующую активность, служить сигнальными молекулами, обеспечивающими взаимодействие *L. brevis* с другими бактериальными и эукариотическими клетками.
Выводы: Использование комплекса омиксных технологий для характеристики перспективных штаммов, может являться перспективным направлением для обнаружения биологически-активных веществ и механизмов их действия. В свою очередь, многие аспекты подобных исследований требуют доработки. В частности – методики стандартизации и сопоставления результатов, полученных при помощи различных омиксных технологий.
Финансирование: Исследование поддержано Российским научным фондом (Грант № 23-16-00123).

Development of algorithms and approaches for the studying promising strains of pharmacobiotic using omics technologies

Marsova M.[1, 2]*, Reznikova D.[1, 2], Letvinava V.[1], Galanova O.[1, 2]

[1] *Vavilov Institute of General Genetics RAS (VIGG), Moscow, Russia*

[2] *Moscow Institute of Physics and Technology (National Research University), (MIPT), Moscow, Russia*

* *masha_marsova@mail.ru*

Key words: omics technologies; pharmacobiotics; proteomic analysis; antioxidants; microbiome

Motivation and Aim: Differences between the microbiomes of clinically healthy people and people with health problems can serve as both a diagnostic sign and a goal for subsequent therapy. In light of the disappointing statistics on the increase in cases of non-communicable chronic diseases, the possibility of microbiome correction takes on new meanings. At the same time, the prospects for the practical application of traditional pre- and probiotics remain blurred. They should be replaced by a new generation of standardized drugs aimed at correcting specific nosologies.

Methods and Algorithms: For the study, the *Levilactobacillus brevis* 47f strain was taken as a promising pharmacobiotic, which showed the greatest adaptogenic and antioxidant activity among related strains in *in vitro* models (using bioluminescent tests) and *in vivo* (on models of stress in mice and low-dose exposure to xenobiotics in *Danio rerio* fish) [1–4].

To identify the active principle of *L. brevis* 47f used both traditional biological models and genomic analysis, as well as omics approaches: proteomic, transcriptomic and metabolomic methods.

The study used the results of comparative genomic analysis of *L. brevis* 47f (WGS PRJNA280953 in the GenBank database) and related strains from the NCBI database, as well as a catalog of genes capable of influencing the manifestation of antioxidant activity [5, https://github.com/Alexey-Kovtun/Catalog].

Total RNA for transcriptomic analysis was isolated at the automated KingFisher station using MagMax mirVana. The finished libraries were sequenced on the Illumina HiSeq 2500 device. To compile a reading map and estimate the number of transcripts, Kallisto software version 0.46.0 was used. Differential expression analysis was performed using edgeR package v3.26.8 integrated into the Degust v4.1.1 web tool.

Intracellular proteins was performed using reverse-phase chromatography on the Ultimate 3000 Nano LC System (Thermo Fisher Scientific) device connected to the Q Exact Plus Orbitrap mass spectrometer (Thermo Fisher Scientific) by means of a nanoelectrospray source (Thermo Fisher Scientific). The obtained data were analyzed using the program Peak studio 10.0 (BioinformaticsSolutions Inc.). Protein identification was performed by correlating tandem mass spectra with a database of protein sequences from the Uniprot database.

Volatile metabolites of the culture fluid were analyzed on a GC device/MS Shimadzu QP2010 Ultra with steam phase extractor Shimadzu HS-20, column VF-WAX MS. To process the chromatogram, Shimadzu LabSolutions software and the NIST14 mass spectrum library with the AMDIS package were used, which detects compounds on the chromatogram and receives purified mass spectra for each compound.

Gel electrophoresis of intracellular proteins and proteins of culture fluid with mass spectrometric examination of characteristic bands was performed as additional methods. The data of proteomic, transcriptomic and metabolomic analysis were analyzed using the developed bioformatic algorithms using the Python programming language and machine learning. The obtained results were compared with the data of standard physico-chemical methods.

Results: Comparative genomic analysis revealed a number of genes uncharacteristic for *Levilactobacillus brevis*, the products of which are able to influence the properties of the strain manifested in biological models. For example, glutaredoxin-like protein NrdH, glutathione reductase/NAD(F)/FAD-dependent oxidoreductase, Dyp-type peroxidase, etc. Transcripts of the corresponding genes, as well as proteins, were found in the results of transcriptomic and proteomic analysis, respectively. A comparison of the results obtained using omics technologies and physico-chemical analysis methods allows us to make assumptions about possible mechanisms and metabolites involved in the implementation of the observed antioxidant activity. In particular, some of the identified proteins and metabolites can exhibit not only antioxidant, but also immunomodulatory activity, serve as signaling molecules that ensure the interaction of *L. brevis* with other bacterial and eukaryotic cells.

Conclusion: The use of a complex of office technologies to characterize promising strains, search for biologically active substances and the mechanisms of their action may be a promising direction. In turn, many aspects of such studies require improvement. In particular, the methods of standardization and comparison of the results obtained with the help of omics technologies.

Funding: The study is supported by Russian Science Foundation (No. 23-16-00123).

Список литературы/References

1. Marsova M., Abilev S., Poluektova E., Danilenko V. A bioluminescent test system reveals valuable antioxidant properties of lactobacillus strains from human microbiota. *World J Microbiol Biotechnol.* 2018;34(2):27. doi 10.1007/s11274-018-2410-2
2. Marsova M., Odorskaya M., Novichkova M. et al. TheLactobacillus brevis 47f strain protects the murine intestine from enteropathy induced by 5-fluorouracil. *Microorganisms*. 2020;8(6):876. doi 10.3390/microorganisms8060876
3. Olekhnovich E.I., Batotsyrenova E.G., Yunes R.A. et al. The effects of Levilactobacillus brevis on the physiological parameters and gut microbiota composition of rats subjected to desynchronosis. *Microb Cell Fact*. 2021;20:226. doi 10.1186/s12934-021-01716-x
4. Kochetkov N., Smorodinskaya S., Vatlin A. et al. Ability of lactobacillus brevis 47f to alleviate the toxic effects of imidacloprid low concentration on the histological parameters and cytokine profile of zebrafish (danio rerio). *Int J Mol Sci*. 2023;24(15):12290. doi 10.3390/ijms241512290
5. Averina O.V., Poluektova E.U., Marsova M.V., Danilenko V.N. Biomarkers and utility of the antioxidant potential of probiotic Lactobacilli and bifidobacteria as representatives of the human gut microbiota. *Biomedicines*. 2021;9(10):1340. doi 10.3390/biomedicines9101340
6. Poluektova E.U., Averina O.V., Kovtun A.S., Danilenko V.N. Transcriptomic analysis of *Levilactobacillus brevis* 47f strain under conditions of oxidative stress. *Genetics.* 2023;59(8):888-897

doi 10.18699/bgrs2024-8.1-08

Поиск новых постбиотиков с использованием омиксных технологий для применения в аквакультуре

Резникова Д.[1, 2]*, Ватлин А.[1], Марсова М.[1], Даниленко В.[1]
[1] *Институт Общей Генетики им. Н.И. Вавилова РАН, Москва, Россия*
[2] *Московский физико-технический институт, Долгопрудный, Россия*
** reznikova.da@phystech.edu*

Ключевые слова: постбиотики; аквакультура; омиксные технологии; *Levilactobacillus brevis* 47f

Мотивация и цель: В результате сельскохозяйственной и промышленной деятельности возрастает загрязнённость водоёмов низкими дозами ксенобиотиков, которые оказывают пагубное влияние на аквакультуру [1]. В последнее время для уменьшения этого эффекта используют пробиотики и, как результат их деятельности, постбиотики.
Выбор штамма *Levilactobacillus brevis* 47f из коллекции лаборатории генетики микроорганизмов основывался на его антиоксидантных и адаптогенных свойствах, доказанных в предыдущих исследованиях [2, 3]. Целью работы было исследование возможности применения культуры как адаптогена при воздействии низких доз ксенобиотиков в аквакультуре, а также выявление потенциальных молекулярных механизмов её действия.
Методы и алгоритмы: В данном исследовании был применен мультиомиксный подход для комплексного решения поставленной задачи. Анализ дифференциальной экспрессии генов был произведён по данным RNAseq, они были проанализированы с использованием специальных программных пакетов, а сама дифференциальная экспрессия была произведена с помощью edgeR v3.42.4. Протеомный анализ белков клеточной фракции производился с помощью хроматографа Ultimate 3000 Nano LC System, соединённого с масс-спектрометром Q Exactive Plus Orbitrap посредством наноэлектроспрейного источника. Анализ протеомных данных производился с использованием Picks Studio 10.0. Метаболомный анализ экспрессируемых компонентов бактериальной клетки производился с помощью GC-MS.
Результаты: В процессе исследования был проведён комплексный анализ штамма, подверженного влиянию различных доз пластификатора бисфенол А. В результате транскриптомного анализа было выявлено изменение экспрессии ряда генов, в том числе генов транспортных белков, генов синтеза жирных кислот, транскрипционных регуляторов, генов *araA, araB, araC*, задействованных в метаболизме арабинозы и гена *oppA*, участвующего в защитных механизмах к различным стрессам. Методом хромато-масс-спектрометрии был проведён протеомный анализ штамма и его продуктов, получено изменение спектра белков на фоне стресса. Примечательным является наличие в спектре таких белков, как L-арабионозоизомераза и субъединица Clp шаперона, которые, имея связь через биохимические пути, предположительно, участвуют в процессах адаптогенности штамма. Метаболомный анализ водной и органической фазы культуральной жидкости штамма способствовал выделению классов метаболитов, которые, вероятно, задействованы в его адаптогенных свойствах.

Выводы: Таким образом, применение омиксных технологий способствовало поиску молекулярных механизмов действия штамма *Levilactobacillus brevis* 47f и выявлению причин его адаптогенных свойств при воздействии различных доз ксенобиотиков, что является неотъемлемой частью при разработке эффективных постбиотиков.
Финансирование: Исследование поддержано грантом Российского научного фонда (№ 23-16-00123).

The search for new postbiotics using omics technologies for aquaculture applications

Reznikova D.[1, 2]*, Vatlin A.[1], Marsova M.[1], Danilenko V.[1]

[1] *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

* *reznikova.da@phystech.edu*

Key words: postbiotics; aquaculture; omics technologies; *Levilactobacillus brevis* 47f

Motivation and Aim: As a result of agricultural and industrial activities, there is increasing pollution of water bodies with low doses of xenobiotics that have detrimental effects on aquaculture [1]. Recently, probiotics and, as a result of their activities, postbiotics have been used to reduce this effect.
The selection of *Levilactobacillus brevis* 47f strain from the collection of laboratory of bacterial genetics was based on its antioxidant and adaptogenic properties proven in previous studies [2, 3]. The aim of the work was to investigate the possibility of using the culture as an adaptogen during exposure to low doses of xenobiotics in aquaculture and to identify potential molecular mechanisms of its action.
Methods and Algorithms: In this study, we applied a multi-omics approach to solve the problem in a comprehensive manner. Differential gene expression analysis was performed using RNAseq data, it was analyzed using special software packages, and differential expression itself was performed using edgeR v3.42.4. Proteomic analysis of cell fraction proteins was performed using an Ultimate 3000 Nano LC System chromatograph coupled to a Q Exactive Plus Orbitrap mass spectrometer via a nanoelectrospray source. Proteomic data were analyzed using Picks Studio 10.0. Metabolomic analysis of expressed bacterial cell components was performed using GC-MS.
Results: In the course of the study, a comprehensive analysis of the strain exposed to different doses of the plasticiser bisphenol A was carried out. Transcriptome analysis revealed changes in the expression of a number of genes, including transport protein genes, fatty acid synthesis genes, transcriptional regulators, *araA, araB, araC* genes involved in arabinose metabolism and *oppA* gene involved in defence mechanisms to various stresses. Proteomic analysis of the strain and its products was carried out by chromatography-mass spectrometry, and the change in the protein spectrum against stress was obtained. Noteworthy is the presence in the spectrum of such proteins as L-arabinose isomerase and Clp chaperone subunit, which are connected through biochemical pathways, presumably participating in the adaptogenicity processes of the strain. Metabolomic analysis of the aqueous and organic phase of the culture fluid of the

strain contributed to the identification of classes of metabolites that are probably involved in its adaptogenic properties.

Conclusion: Thus, the application of omics technologies contributed to the search for molecular mechanisms of action of *Levilactobacillus brevis* 47f strain and identification of the causes of its adaptogenic properties when exposed to different doses of xenobiotics, which is an integral part in the design of effective postbiotics.

Funding: The study is supported by a grant from the Russian Science Foundation (No. 23-16-00123).

Список литературы/References

1. Rohani M.F. Pesticides toxicity in fish: Histopathological and hemato-biochemical aspects – A review. *Emerging Contam.* 2023;9(3):100234
2. Marsova M. et al. A bioluminescent test system reveals valuable antioxidant properties of lactobacillus strains from human microbiota. *World J Microbiol Biotechnol*. 2018;34(2):27. doi 10.1007/s11274-018-2410-2
3. Olekhnovich E.I. et al. The effects of Levilactobacillus brevis on the physiological parameters and gut microbiota composition of rats subjected to desynchronosis. *Microb Cell Fact.* 2021;20(1):226

doi 10.18699/bgrs2024-8.1-09

Эффекты применения пробиотического штамма U-21 в комбинированной модели паркинсонического синдрома у крыс

Ставровская А.[1]*, Воронков Д.[1], Ольшанский А.[1], Кутукова К.[1], Марсова М.[2], Даниленко В.[2], Иллариошкин С.[1]

[1] *ФГБНУ «Научный центр неврологии», Москва, Россия*

[2] *ФГБУН Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия*

** alla_stav@mail.ru*

Ключевые слова: болезнь Паркинсона; нейровоспаление; пробиотик; биомодель

Мотивация и цель: Данные об участии микробиоты в развитии болезни Паркинсона позволяют обсуждать способность бактериальных препаратов влиять на процессы, приводящие к нейродегенерации [1]. Целью работы была оценка влияния фармабиотика U-21 на основе штамма Limosilactobacillus fermentum на исходы модели БП, полученной при комбинированном воздействии ЛПС и параквата.

Методы и алгоритмы: Работа проведена на 30 самцах крыс Вистар (вес 330 г). Животным (n = 18) в компактную часть черной субстанции (кЧС) [2] справа вводили раствор липополисахарида из E.Coli (LPS) (4 мкг в 3 мкл физраствора), с левой стороны – тот же объем растворителя. Для анестезии применяли ксило-золетиловый наркоз. Начиная со 2-го дня после операции крысам через день внутрибрюшинно провели 7 инъекций параквата (Pq) в дозе 8 мг/кг в 0.9% NaCl. Контрольным животным (n = 12) проводили билатеральные интранигральные и внутрибрюшинные инъекции физраствора. Фармабиотик U-21 (108 КОЕ) [3] разводили в 3 мл физиологического раствора и вводили перорально, ежедневно по 0.3 мл раствора.

Далее животные были разделены на 4 группы: “NaCl+NaCl+NaCl“, “NaCl+NaCl+U-21“ (по 6 в каждой группе) и “LPS+Pq+NaCl“ и “LPS+Pq+U-21“ (по 9 в каждой группе). Нарушения локомоции оценивали в тесте «Сужающаяся дорожка» (СД) [4] по количеству соскальзываний конечностей на нижнюю планку дорожки (в % от общего количества шагов). Тест проводили по окончании введения препаратов.

Для морфологического исследования животных декапитировали гильотиной, мозг извлекали и фиксировали 24 ч в 10 % формалине. На криостатных срезах ЧС (10 мкм), после обработки в цитратном буфере (pH = 6.0, 15 мин при 95 °C) иммунофлуоресцентным методом по протоколам производителей антител выявляли: тирозингидроксилазу (TH), α-синуклеин, фосфорилированный по остатку Ser-129, компонент C3 комплемента, нейрональный белок NeuN, глиофибриллярный белок астроцитов GFAP и маркер микроглии IBA1. Для визуализации применяли

антитела с флуорохромами CF488 и CF555, срезы заключали в среду Fluoroshield с DAPI.

Используя микроскопы Nikon Eclipse Ni-u и Nikon SMZ-25 и программу NIS Elements BR 4.2, на сериях из 12–18 срезов, взятых с равным интервалом на всем протяжении ЧС, оценивали интенсивность флуоресценции (в градациях яркости, 10 бит). Результаты для каждого животного усредняли. Вычисляли изменения в процентах, относительно левой, контрольной ЧС (без введения LPS) и обрабатывали в программе Prism 8.0, используя дисперсионный анализ (ANOVA) и тест Тьюки. Различия считали статистически значимыми при $p < 0.05$.

При иммунофлуоресцентном исследовании кишечника образцы тощей кишки фиксировали в течение 1 суток забуференным формалином, пропитывали 30 % раствором сахарозы и на криостате получали срезы толщиной 10 мкм. Иммунофлуоресцентные реакции проводили непрямым методом, используя первичные антитела к общему α-синуклеину (α-Syn), α-синуклеину, фосфорилированному по серину-129 (α-Syn-pS129) и маркер апоптоза – активированную каспазу 3, а также вторичные антитела, меченные флуорохромами CF448 и Cy3. Препараты исследовали и фотографировали под микроскопом NikonEclipse NiU с цифровой камерой Nikon DS-Qi. Морфометрию выполняли с помощью программы ImageJ (30–40 полей зрения на животное). Содержание общего и фосфорилированного α-синуклеина в миентеральных ганглиях оценивали косвенно путём измерения интенсивности флуоресцентного окрашивания на эти маркеры на площадь ганглия. Статистический анализ проводили в программе Statistica 8.0. при помощи критерия Манна–Уитни, различия признавали статистически значимыми при $p < 0.05$.

Результаты: Введение токсинов привело к увеличению числа соскальзываний задними конечностями, как по сравнению с числом соскальзываний передних конечностей у животных той же группы ($p = 0.021$), так и с числом соскальзываний задних конечностей по сравнению с контрольными группами ($p = 0.004$; $p = 0.018$). При введении фармабиотика U-21 была выявлена тенденция к сглаживанию выявленных нарушений локомоции как между модельными, так и контрольными группами.

Средняя интенсивность окрашивания на ТН, на стороне введения LPS значимо ($p = 0.004$) снижалась в обеих группах с моделированием БП (“LPS+Pq+U-21” и “LPS+Pq+NaCl”) относительно контрольных групп.

Статистически значимое ($p = 0.012$) увеличение интенсивности окрашивания на α-Syn-pS129 выявили на стороне введения LPS в группе, не получавшей U-21. Содержащих α-синуклеин агрегатов у животных, получавших LPS и Pq, не выявили, но обнаруживали интенсивное окрашивание на α-Syn-pS129 в не-дофаминергических нервных волокнах и округлые, негативные к тирозингидроксилазе, внеклеточные включения.

В ЧС животных, получавших LPS+Pq, отмечали реактивные изменения IBA1-позитивной микроглии и GFAP-содержащих астроцитов относительно контроля. В группе ”LPS+Pq+NaCl” деформация и гипертрофия отростков глии и усиление окрашивания были выраженнее, чем в группе ”LPS+Pq+U-21”.
Окрашивание на C3 компонент комплемента в ЧС у контрольных животных выявляли преимущественно в нейропиле, тогда как в обеих группах, получавших LPS+Pq, отмечали окрашивание тел нейронов. При этом выраженность провоспалительной реакции в группе ”LPS+Pq+NaCl” была выше, чем в других группах, что подтверждалось увеличением интенсивности окрашивания на компонент C3 комплемента ($p = 0.047$).
Исследование образцов кишечника показало, что введение LPS+Pq не вызывало апоптоза в ганглиях межмышечного сплетения тонкой и толстой кишки. Дополнительное введение штамма заметно снижало уровень апоптоза клеток эпителия толстой кишки. Кроме этого, введение LPS+Pq вызывало повышение интенсивности иммунофлуоресцентного окрашивания на α-Syn-pS129 в миентеральных ганглиях тонкой кишки, по сравнению с контролем. Введение штамма крысам-моделям БП достоверно снижало интенсивность иммунофлуоресцентного окрашивания на α-Syn-pS129 в миентеральных ганглиях тонкой и толстой кишки.
Выводы: Применение фармабиотика не привело к значимому ослаблению нарушений локомоции крыс-биомоделей. Морфологическое исследование показало, что штамм U-21 не повлиял на степень повреждения нейронов ЧС в комбинированной модели LPS+Pq, но при этом использование фармабиотика снижало накопление α-Syn-pS129 и провоспалительные изменения в веществе мозга. В целом, полученные результаты свидетельствуют о явном потенциале фармбиотика U-21 для предотвращения накопления БП-ассоциированного α-Syn иα-Syn-pS129, а также его возможного дальнейшего распространения по структурам ПНС и ЦНС на начальных стадиях развития паркинсонизма.
Финансирование: Работа поддержана грантом Министерства науки и высшего образования РФ для проведения крупных научных проектов по приоритетным направлениям научно-технологического развития (проект № 075-15-2024-638).

Effects of the probiotic strain U-21 in a combined model of parkinsonian syndrome in rats

Stavrovskaya A.[1]*, Voronkov D.[1], Olshansky A.[1], Kutukova K.[1], Marsova M.[2], Danilenko V.[2], Illarioshkin S.[1]

[1] *Research Center of Neurology, Moscow, Russia*

[2] *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia*

* *alla_stav@mail.ru*

Key words: Parkinson's disease; neuroinflammation; probiotic; biomodel

Motivation and Aim: Data on the involvement of microbiota in the development of Parkinson's disease allow discussing the ability of bacterial preparations to influence the processes leading to neurodegeneration [1]. The aim of the work was to evaluate the effect of the pharmabiotic U-21 based on the Limosilactobacillus fermentum strain on the outcomes of a PD model obtained by the combined action of LPS and paraquat.

Methods and Algorithms: The work was carried out on 30 male Wistar rats (weight 330 g). The animals (n = 18) were injected with a solution of lipopolysaccharide from E. Coli (LPS) (4 μg in 3 μl of saline) into the compact part of the substantia nigra (SNc) [2] on the right, and the same volume of solvent on the left side. Xyla-zoletil anesthesia was used. Starting from the 2nd day after the operation, the rats were injected intraperitoneally with paraquat (Pq) at a dose of 8 mg/kg in 0.9 % NaCl every other day for 7 injections. Control animals (n = 12) underwent bilateral intranigral and intraperitoneal injections of saline. The pharmabiotic U-21 (108 CFU) [3] was diluted in 3 ml of saline and administered orally, daily at 0.3 ml of solution.

The animals were then divided into 4 groups: "NaCl+NaCl+NaCl", "NaCl+NaCl+U-21" (6 in each group) and "LPS+Pq+NaCl" and "LPS+Pq+U-21" (9 in each group). Locomotor disorders were assessed in the "Narrowing Beam" test (NB) [4] by the number of slips of the limbs onto the lower plank of the track (as a % of the total number of steps). The test was carried out at the end of drug administration.

For morphological examination, the animals were decapitated with a guillotine, the brain was removed and fixed for 24 h in 10 % formalin. On cryostat sections of the SNc (10 μm), after treatment in citrate buffer (pH = 6.0, 15 min at 95 °C), the following were detected by immunofluorescence according to the protocols of the antibody manufacturers: tyrosine hydroxylase (TH), α-synuclein phosphorylated at Ser-129 (α-Syn-pS129) residue, complement component C3, neuronal protein NeuN, glial fibrillary acidic protein GFAP and microglia marker IBA1. For visualization, antibodies with CF488 and CF555 fluorochromes were used, the sections were enclosed in Fluoroshield medium with DAPI.

Using Nikon Eclipse Ni-u and Nikon SMZ-25 microscopes, and the NIS Elements BR 4.2 program, on a series of 12–18 sections taken at equal intervals throughout the SNc, the fluorescence intensity was assessed (in brightness gradations, 10 bits). The results for each animal were averaged. Changes were calculated as a percentage relative to the left, control SNc (without LPS injection) and processed in the Prism 8.0 program using analysis of variance (ANOVA) and Tukey's test. Differences were considered statistically significant at $p < 0.05$.

In the immunofluorescence study of the intestine, samples of the jejunum were fixed for 1 day with buffered formalin, impregnated with a 30 % sucrose solution, and 10 μm thick sections were obtained on a cryostat. Immunofluorescence reactions were carried out by the indirect method using primary antibodies to total α-synuclein (α-Syn), α-Syn-pS129 and the apoptosis marker activated caspase 3, as well as secondary antibodies labeled with CF448 and Cy3

fluorochromes. The preparations were examined and photographed under a Nikon Eclipse NiU microscope with a Nikon DS-Qi digital camera. Morphometry was performed using the ImageJ program (30–40 fields of view per animal). The content of total and phosphorylated α-synuclein in the myenteric ganglia was assessed indirectly by measuring the intensity of fluorescent staining for these markers per ganglion area. Statistical analysis was performed in the Statistica 8.0 program using the Mann–Whitney test, differences were considered statistically significant at $p < 0.05$.

Results: The administration of toxins led to an increase in the number of slips by the hind limbs, both in comparison with the number of slips of the forelimbs in animals of the same group ($p = 0.021$), and with the number of slips of the hind limbs compared to the control groups ($p = 0.004$; $p = 0.018$). With the administration of the pharmabiotic U-21, a tendency to smooth out the identified locomotor disorders was revealed both between the model and control groups.

The average staining intensity for TH on the side of LPS injection significantly ($p = 0.004$) decreased in both groups with PD modeling ("LPS+Pq+U-21" and "LPS+Pq+NaCl") relative to the control groups.

A statistically significant (p=0.012) increase in the staining intensity for α-Syn-pS129 was revealed on the side of LPS injection in the group not receiving U-21. Aggregates containing α-synuclein were not detected in animals receiving LPS and Pq, but intense staining for α-Syn-pS129 was found in non-dopaminergic nerve fibers and round, tyrosine hydroxylase-negative, extracellular inclusions.

In the SNc of animals receiving LPS+Pq, reactive changes in IBA1-positive microglia and GFAP-containing astrocytes were noted relative to the control. In the "LPS+Pq+NaCl" group, the deformation and hypertrophy of glial processes and increased staining were more pronounced than in the "LPS+Pq+U-21" group.

Staining for complement component C3 in the SNc in control animals was detected mainly in the neuropil, whereas in both groups receiving LPS+Pq, staining of neuronal bodies was noted. At the same time, the severity of the pro-inflammatory reaction in the "LPS+Pq+NaCl" group was higher than in other groups, which was confirmed by an increase in the staining intensity for complement component C3 ($p = 0.047$).

The study of intestinal samples showed that the introduction of LPS+Pq did not cause apoptosis in the ganglia of the intermuscular plexus of the small and large intestine. Additional administration of the strain noticeably reduced the level of apoptosis of colon epithelial cells. In addition, the introduction of LPS+Pq caused an increase in the intensity of immunofluorescent staining for α-Syn-pS129 in the myenteric ganglia of the small intestine, compared with the control. The introduction of the strain to PD model rats significantly reduced the intensity of immunofluorescent staining for α-Syn-pS129 in the myenteric ganglia of the small and large intestine.

Conclusion: The use of the pharmabiotic did not lead to a significant reduction in locomotor disorders in bio-model rats. Morphological study showed that the U-21 strain did not affect the degree of damage to SNc neurons in the combined

LPS+Pq model, but at the same time, the use of the pharmabiotic reduced the accumulation of α-Syn-pS129 and pro-inflammatory changes in the brain substance. In general, the results obtained indicate the clear potential of the U-21 pharmabiotic for preventing the accumulation of PD-associated α-Syn and α-Syn-pS129, as well as its possible further spread through the structures of the PNS and CNS at the initial stages of the development of parkinsonism.

Funding: The study was supported by a grant from the Ministry of Science and Higher Education of the Russian Federation for major scientific projects in priority areas of scientific and technological development (project No. 075-15-2024-638).

Список литературы/References

1. Alipour Nosrani E., Tamtaji O.R., Alibolandi Z. et al. Neuroprotective effects of probiotics bacteria on animal model of Parkinson's disease induced by 6-hydroxydopamine: A behavioral, biochemical, and histological study. *J Immunoassay Immunochem*. 2021;42(2):106-120. doi 10.1080/15321819.2020.1833917
2. Иллариошкин С.Н., Левин О.С., Федотова Е.Ю., Колоколов О.В. Паркинсонизм и Черная Субстанция. Судьба Великого Открытия. М.: ООО Издательское предприятие “Атмосфера”, 2019
3. Даниленко В.Н., Марсова М.В., Полуэктова Е.У. и др. Штамм Lactobacillus fermentum U-21, продуцирующий комплекс биологически активных веществ, осуществляющих нейтрализацию супероксид-аниона, индуцируемого химическими агентами. Патент на изобретение RU 2705250 C2, 06.11.2019. Заявка № 2018104322 от 05.02.2018
4. Sweis B.M., Bachour S.P., Brekke J.A. et al. A modified beam-walking apparatus for assessment of anxiety in a rodent model of blast traumatic brain injury. *Behav Brain Res*. 2016;296:149-156. doi 10.1016/j.bbr.2015.09.015

doi 10.18699/bgrs2024-8.1-10

Исследование разнообразия и противомикробного потенциала кишечного вирома здоровых людей и пациентов с язвенным колитом

Федорец В.*, Тикунов А., Матвеев А., Тикунова Н.
Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** v.fedorets@g.nsu.ru*

Ключевые слова: кишечный виром; язвенный колит; метагеном; эндолизины

Мотивация и цель: Известно, что виром кишечника человека представлен преимущественно бактериофагами, которые играют важную роль в формировании микробных сообществ, регулируя численность своих бактериальных хозяев [1]. В связи с этим все более широкую распространенность приобретают исследования, направленные на сравнение вирома здоровых людей и пациентов с различными хроническими заболеваниями кишечника, такими как язвенный колит (ЯК). Кроме того, в последнее время все острее встает проблема растущей антибиотикорезистентности бактерий, поэтому актуальной также является задача поиска новых противомикробных средств. Такими антибактериальными агентами могут являться эндолизины – ферменты бактериофагов, используемые вирусами для разрушения клеточной стенки бактерии. Целью данной работы было исследование кишечного вирома здоровых людей и пациентов с ЯК, а также поиск в нем генов потенциальных эндолизинов и оценка их разнообразия.
Методы и алгоритмы: Из 17 образцов стула (8 от здоровых добровольцев (ЗД) и 9 от пациентов с ЯК) была отделена обогащенная вирусными частицами фракция, из которой в дальнейшем была выделена и секвенирована тотальная ДНК. Полученные прочтения были отфильтрованы по качеству с помощью Trimmomatic и собраны *de novo* в контиги с помощью SPAdes. Потенциальные геномы бактериофагов и их фрагменты были выявлены с помощью программного обеспечения VIBRANT v.1.2.1 [2]. Статистическую обработку результатов, полученных при сравнении выборок, производили с помощью критерия Манна–Уитни. Также с помощью VIBRANT были определены (с использованием Prodigal) и аннотированы (с использованием баз данных KEGG, Pfam и VOG) открытые рамки считывания, содержащиеся в выявленных потенциальных геномах. На основании полученных аннотаций была собрана база данных, содержащая аминокислотные последовательности потенциальных эндолизинов. Затем для фаговых последовательностей, содержащих гены эндолизинов, осуществляли прогнозирование бактериальных хозяев с помощью интегрированной среды машинного обучения iPHoP [3], которая объединяет различные инструменты, используемые для решения этой задачи: поиск сходства между геномом вируса и геномом хозяина, анализ CRISPR кассет, анализ частот встречаемости k-меров, а также сравнение с базой данных уже изученных бактериофагов с известными хозяевами.
Результаты: В виромах, полученных из образцов от ЗД, медианное количество выявленных фаговых последовательностей составило 71, при этом распределение

по образцам было не равномерным – от 10 до 261. В образцах от пациентов с ЯК медианное количество фаговых последовательностей составило 115 (от 13 до 195), что оказалось достоверно выше, чем для ЗД (p-value < 0.05). Кроме того, количество выявленных полных кольцевых геномов бактериофагов было также достоверно выше в выборке пациентов с ЯК, чем ЗД (медианное значение 0.5 для ЗД, 5 для пациентов с ЯК, p-value < 0.05), что может объясняться более высоким покрытием, полученным в результате сборки контигов. На основании полученных с помощью VIBRANT аннотаций всего в виромах было выявлено 134 последовательности потенциальных эндолизинов, 85 из которых – у пациентов с ЯК, 49 – у ЗД. Что касается предположительных бактериальных хозяев потенциальных бактериофагов, полные или частичные последовательности которых содержали гены эндолизинов, для выборки ЗД наиболее распространенными оказались представители филума Bacteroidota, класса Bacteroidia, которые составили 54.5 % от общего количества хозяев, определенных с высокой степенью достоверности. В основном это были бактерии родов *Bacteroides*, *Phocaeicola*, *Prevotella*. Вторым распространенным таксоном бактерий оказался класс Clostridia, относящийся к филуму Bacillota (39.5 %), представленный большим количеством родов. Стоит отметить, что представители двух данных классов составляют подавляющую часть здоровой микробиоты кишечника человека. Что касается пациентов с ЯК, наиболее распространенными хозяевами бактериофагов у них оказались также представители классов Bacteroidia и Clostridia (36.4 и 34.5 % соответственно), однако обращает на себя внимание снижение относительной доли первого. При этом у пациентов были также выявлены потенциальные фаги, заражающие бактерии классов Actinomicetia (а именно род *Bifidobacterium*) и Gammaproteobacteria. Что касается структуры выявленных эндолизинов, наиболее часто встречающимися каталитическими доменами были *Glyco_hydro_25* и *Glucosaminidase*, которые более характерны для эндолизинов фагов грамположительных бактерий, а также *Phage_lysozyme* и *Muraidase*, более характерные для эндолизинов, воздействующих на грамотрицательные бактерии. Также часто встречающимися были домены *Amidase_2* и *Peptidase_M23*, выявляющиеся в эндолизинах как грамположительных, так и грамотрицательных бактерий. Среди связывающих доменов наиболее распространенным был *LysM*. Выравнивание с помощью BLASTP показало, что значительная часть выявленных белков имеет сходство менее 50 % с уже охарактеризованными эндолизинами.

Выводы: Сравнительный анализ кишечных виромов показал, что общее количество выявленных потенциальных фаговых последовательностей у пациентов с ЯК достоверно выше, чем у ЗД, что подтверждает предположение о изменении состава вирусных сообществ кишечника при хронических воспалительных заболеваниях. Также у пациентов оказался шире спектр потенциальных бактериальных хозяев, что может быть как причиной, так и следствием нарушений в составе микробиоты у людей с ЯК. Кроме того, в исследуемых виромных данных были обнаружены последовательности 134 потенциальных эндолизинов с разнообразной доменной структурой, многие из которых не имеют охарактеризованных гомологов, и их дальнейшее исследование может являться ступенью к разработке новых противомикробных средств.

Финансирование: Исследование поддержано Госзаданием 0245-2022-0005 (№ 075-03-2022-371/4) ИХБФМ СО РАН.

The gut virome of healthy people and patients with ulcerative colitis: diversity and antimicrobial potential

Fedorets V.*, Tikunov A., Matveev A., Tikunova N.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** v.fedorets@g.nsu.ru*

Key words: gut virome; ulcerative colitis; metagenome; endolysins

Motivation and Aim: It is known that human intestinal virome is represented mainly by bacteriophages, which play an important role in the microbiota formation via regulating the number of their bacterial hosts [1]. In this regard, studies aimed at comparing the virome of healthy people and patients with various chronic intestinal diseases, such as ulcerative colitis (UC), are becoming increasingly common. Additionally, the problem of growing bacterial antibiotic resistance has become increasingly acute, so the issue of searching for new antimicrobial agents is also urgent. Such antibacterial agents can be endolysins, enzymes used by phages to destroy the bacterial cell wall. The purpose of this work was to study the intestinal virome of healthy people and patients with UC, as well as to search for potential endolysin genes and assess their diversity.
Methods and Algorithms: Stool samples from 8 healthy volunteers (HV) and 9 patients with UC were used to separate a fraction enriched in viral particles, from which total DNA was subsequently isolated and sequenced. The resulting reads were quality filtered using Trimmomatic and assembled *de novo* into contigs using SPAdes. Potential bacteriophage genomes and their fragments were identified using VIBRANT v.1.2.1 software [2]. Statistical processing of the results was carried out using the Mann–Whitney test. VIBRANT was also used to identify (via Prodigal) and annotate (via KEGG, Pfam and VOG databases) the open reading frames in the candidate genomes. Based on the annotations obtained, a database containing the amino acid sequences of putative endolysins was compiled. Then, for potential phage sequences containing endolysin genes, bacterial hosts were predicted using the integrated machine learning framework iPHoP [3], which combines various tools used to solve this problem: searching for similarities between the viral genome and the host genome, analysis of CRISPR cassettes, analysis of the k-mers occurrence frequency, as well as comparison with a database of already studied bacteriophages with known hosts.
Results: In viromes from HV, the median number of identified phage sequences was 71, while the distribution across samples was uneven – from 10 to 261. In samples from patients with UC, the median number of phage sequences was 115 (from 13 to 195), which turned out to be significantly higher than for HV (p-value < 0.05). In addition, the number of identified complete circular genomes of bacteriophages was also significantly higher in the sample of patients with UC than in HV (median value 0.5 for HV, 5 for patients with UC, p-value < 0.05), which may be explained by a higher coverage after contig assembly. Based on the annotations obtained using VIBRANT, a total of 134 potential endolysins sequences were identified in the viromes, 85 of them were from patients with UC, 49 from HV. As for the putative bacterial hosts of the identified potential bacteriophages, the full or partial sequences of which contained endolysin

genes, for the HV sample, the most common were representatives of the phylum Bacteroidota, class Bacteroidia, which accounted for 54.5 % of the total number of hosts identified with a high degree of confidence. These were mainly bacteria of the genera *Bacteroides*, *Phocaeicola*, and *Prevotella*. The second common taxon of bacteria was the class Clostridia, belonging to the phylum Bacillota (39.5 %), represented by a large number of genera. It is worth noting that representatives of these two classes make up the vast majority of the healthy human intestinal microbiota. As for patients with UC, the most common hosts of bacteriophages were also representatives of the classes Bacteroidia and Clostridia (36.4 and 34.5 %, respectively), however, a decrease in the relative proportion of the former was seen. At the same time, potential phages that infect bacteria of the classes Actinonomicetia (namely the genus *Bifidobacterium*) and Gammaproteobacteria were also identified in patients. Regarding the structure of the identified endolysins, the most frequently occurring catalytic domains were *Glyco_hydro_25* and *Glucosaminidase*, which are more common in phages of Gram-positive bacteria, as well as *Phage_lysozyme* and *Muraidase*, more common in endolysins affecting Gram-negative bacteria. Also frequently occurring were the *Amidase_2* and *Peptidase_M23* domains, found in endolysins of both Gram-positive and Gram-negative bacteria. Among the binding domains, *LysM* was the most abundant. Alignment using BLASTP showed that a significant proportion of the identified proteins have less than 50 % similarity to already characterized endolysins.

Conclusion: A comparative analysis of intestinal viromes showed that the total number of identified potential phage sequences in patients with UC was significantly higher than in HV, which confirms the assumption that the composition of intestinal viral communities changes in chronic inflammatory diseases. Patients also had a wider range of potential bacterial hosts, which may be both a cause and a consequence of the microbiota disturbances in people with UC. In addition, the sequences of 134 potential endolysins with diverse domain structures were discovered in the studied virome data, many of which do not have characterized homologues, and their further study may be a step towards the development of new antimicrobial agents.

Funding: The study is supported by State Assignment 0245-2022-0005 (No. 075-03-2022-371/4) of the ICBFM SB RAS.

Список литературы/References

1. Ogilvie L.A., Jones B.V. The human gut virome: A multifaceted majority. *Front Microbiol*. 2015;1(6):152433
2. Kieft K., Zhou Z., Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*. 2020;8(1):90. doi 10.1186/s40168-020-00867-0
3. Roux S. et al. iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol*. 2023;21(4):e3002083. doi 10.1371/journal.pbio.3002083

doi 10.18699/bgrs2024-8.1-11

Функциональная метаболомика при воспалительных заболеваниях кишечника

Шагалеева О.Ю., Кашатникова Д.А., Воробьева Е.А., Кардонский Д.А., Силантьев А.С., Ефимов Б.А., Иванов В.А., Беспятых Ю.А., Калачнюк Т.Н., Олехнович Е.И., Захаржевская Н.Б.*

ФГБУ Федеральный научно-клинический центр физико-химической медицины имени академика Ю.М. Лопухина Федерального медико-биологического агентства, Москва, Россия

** natazaha@gmail.com*

Ключевые слова: воспалительные заболевания кишечника; ГХ-МС с парофазным способом экстракции; Bacteroides, микробиота

Введение: Микробиота желудочно-кишечного тракта характеризуется широким спектром секреторной активности. При развитии воспалительных заболеваний происходят качественные модификации микробиома, что в значительной степени отражается на компонентном составе летучего метаболома. Функциональная метаболомика способствует оценке динамического состояния микробиоты в норме и патологии, используя для формирования комплексного описания состояния микробиоты данные метагеномного и метаболомного анализа. Если для оценки метагенома используют стандартные методы секвенирования, то для исследования летучего спектра метаболома рассматривают различные методические подходы. В данной работе для оценки летучего спектра метаболитов микробиоты использован метод ГХ/МС с парофазным способом экстракции. Для оценки динамического состояния микробиоты проведены эксперименты по формированию колита у животных (крысы/мыши) с последующей терапией воспаления посредством везикул *Bacteroides fragilis (JIM),* обладающих иммуномодулирующими свойствами.

Методы и алгоритмы: Для формирования острого колита и последующей оценке спектра летучего метаболома на стадиях острого колита и стадии самопроизвольной реконвалесценции использовали крыс-самок линии Wistar из Питомника лабораторных животных (Пущино РАН, Московская область) в возрасте 2 мес. (200–230 г). Животные были случайным образом разделены на три группы: контрольная группа (Контроль) (n = 5); группа опытная, получавшая декстран сульфат натрия в течение 7 дней, с последующей эвтаназией (ДСН7) (n = 5); группа, получавшая декстран сульфат натрия с последующим недельным восстановление без ДСН (ДНС14) (n = 5). Исследование терапевтической эффективности везикул *Bacteroides fragilis (JIM)* посредством комплексного метаболомного и метагеномного анализа в модели острого колита у животных проводилось на мышах линии C57BL/6, возрастом 7–8 недель. Мыши были разделенные на три группы: к(n = 10), ДСН(n = 10), ДСН+В(n = 10). Эксперимент включал в себя два цикла воспаление/ремиссия, каждый по 10 дней. Группа ДСН+В с 11 по 20 дни получала везикулы концентрацией 1 мг/кг. Массу тела, консистенцию стула и наличие крови в стуле регистрировали каждые 3 дня. Гистологическую пробоподготовку проводили согласно методу, описанному в публикациях [1]. Клинические проявления воспаления оценивали с помощью индекса активности заболевания (ИАЗ). Анализ метаболома проводили в образцах стула методом ГХ-МС на приборе

Shimadzu QP2010 Ultra с парофазным экстрактором Shimadzu HS-20 [2]. Данные ГХ/МС обрабатывали с использованием программного обеспечения MetaboAnalyst 5.0 (http://www.metaboanalyst.ca) и GraphPad Prism 8.0.1. Секвенирование гена 16S рРНК для групп экспериментальных мышей было выполнено на платформе MinION™. Статистическое сравнение групп проводилось с использованием непараметрического критерия Манна–Уитни и поправки на множественное сравнение FDR. Расчеты проводились на языке программирования Python, с помощью модулей stats (библиотека scipy) и stats.multitest (библиотека statsmodels). Построение графиков осуществлялось с помощью библиотек seaborn и matplotlib.

Результаты: согласно полученным данным общий балл ИАЗ включал снижение массы тела, консистенцию стула и ректальное кровотечение. Показатель ИАЗ в контрольной группе у крыс оставался близким к 0 на протяжении всех экспериментальных процедур. ИАЗ в группах крыс, получавших ДСН (ДСН7 и ДСН14), был значительно выше по сравнению с контрольной группой и увеличивался на 3-й день. Максимальный балл, рассчитанный в группах крыс, получавших ДСН (ДСН7 и ДСН14), был получен на 7-й день, когда была достигнута острая фаза воспаления. Потерю веса животных, диарею и видимую кровь в стуле наблюдали на 7-й день в обеих группах. По данным гистопатологического исследования, толстая кишка контрольных животных имела неповрежденную слизистую оболочку, тогда как в толстой кишке, групп крыс ДСН7 и ДСН14, наблюдалась воспалительная клеточная инфильтрация, истончение слизистой и подслизистой оболочки и эрозия эпителия. Как и ожидалось, значения ИАЗ снизились в группе ДСН14 после прекращения введения ДСН на 8–14 день. Была явно отмечена способность кишечника к восстановлению на 14-е сутки, когда наблюдалась нормализация консистенции стула и восстановление веса животных. Определенное улучшение гистологической картины наблюдалось в группе ДСН14, в которой увеличивалось количество крипт и бокаловидных клеток, а структура слизистой оболочки была более дифференцированной, однако минимальная воспалительная инфильтрация сохранялась. Согласно полученным ГХ/МС данным изменения метаболомного профиля можно было наблюдать уже на третий день после введения ДСН. При этом основные симптомы колита у крыс, такие как диарея и кровь в стуле, за этот период времени еще не были обнаружены. Острая фаза колита у крыс достигала пика на 7-е сутки после введения ДСН и характеризовалась изменениями стула. Различия в метаболомном профиле также регистрировались на 7-й день. Основные изменения коснулись таких соединений, как уксусная кислота, пропановая кислота, пропановая кислота-2-метил-, бутановая кислота, гексановая кислота. Несмотря на улучшение гистологической картины после прекращения введения ДСН, наблюдалась тенденция к прогрессированию изменений профиля ЛОС. Метаболомные изменения продолжали увеличиваться к 14 дню. Произошли значительные изменения в метаболомном профиле для таких соединений, как уксусная кислота, бутановая кислота, пентановая кислота, пропановая кислота, 2-метил-пропановая кислота, гексановая кислота, зарегистрированная в группе DSS14. Изменения относительного количества метаболитов, таких как бензолпропановая кислота и индол, были отмечены на 11-й день.

Во втором эксперименте предварительно были выделены везикул *Bacteroides fragilis*, которые характеризовались с помощью ПЭМ. Полученные везикулы

использовали для лечения мышей после воздействия ДСН. Всем трем группам на 10-е и 20-е сутки эксперимента проводилось гистологическое исследование тканей кишечника. По данным гистологического исследования в контрольной группе не наблюдались изменения, тогда как в группе ДСН структура слизистой дистальных отделов содержала протяженные очаги воспалительного поражения. В группе ДСН + В наблюдалась меньшая площадь поражения, а повреждения носили очаговый характер и имели значительно меньшую продолжительность. ИАЗ групп ДСН и ДСН+В были значительно выше, чем у контрольной группы, однако в ходе терапии везикулами ИАЗ в группе ДСН+В значительно снизился, по сравнению с группой ДСН. Среди идентифицированных в ходе метаболомного исследования соединений были обнаружены жирные кислоты с короткой, средней и длинной цепью и производные аминокислот. На 10-й день эксперимента между группами ДСН и ДСН+В не было обнаружено существенных различий. На 20-й день, соответствующий окончанию обработки везикулами, можно было выделить три отдельные группы. Контрольная группа значительно отличалась от групп ДСН и ДСН+В по метаболомному спектру. При лечении везикулами в группе ДСН+В можно было наблюдать тенденцию к восстановлению общего профиля метаболитов к контрольным значениям. Анализ изменчивости микробиома в образцах стула на 10й день эксперимента продемонстрировал снижение численности отдельных видов бактерий по сравнению с контрольной группой. К 20-му дню наблюдалась значительная разница между группами ДСН и контрольной группой при этом относительная численность изученных родов бактерий в группе ДСН+В к двадцатому дню наблюдения достигла исходных значений, определенных в контрольной группе. Корреляционный анализ позволил установить, что увеличение Saccharibacteria и Acetivibrio положительно коррелировало с фенолом и пентановой кислотой, а в группе ДСН преобладание Lactococcus и Romboutsia отрицательно коррелировало с фенолом и пентановой кислотой.

Выводы: в модельных экспериментах видно, что симбиоз двух методик (ГХ/МС и метагеномного секвенирования) позволяет неинвазивно отслеживать все этапы воспаления и процессы восстановления эпителия кишечника у подопытных животных. Таким образом данные функциональной метаболомики могут быть использованы для контроля активности воспаления у пациентов с ВЗК.

Functional metabolomics of inflammatory bowel diseases

Shagaleeva O.Yu., Kashatnikova D.A., Vorobyova E.A., Kardonsky D.A., Silantev A.S., Efimov B.A., Ivanov V.A., Bespyatykh Yu.A., Kalachnyuk T.N., Olekhnovich E.I., Zakharzhevskaya N.B.*

Lopukhin Federal Research and Clinical Centre of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia

** natazaha@gmail.com*

Key words: inflammatory bowel diseases; GC-MS with headspace extraction; Bacteroides; microbiota

Motivation and Aim: The microbiota is characterized by a wide range of secretory activities. Qualitative modifications of the microbiome occur during inflammatory diseases, which is largely reflected in the composition of the volatile metabolome.

Functional metabolomics contributes to the assessment of the dynamic state of the microbiota in health and disease, using data from metagenomic and metabolomic analysis to form a comprehensive description of the state of the microbiota. If standard sequencing methods are used to assess the meta-genome, then various methodological approaches are considered to study the volatile spectrum of the metabolome. In this research, the GC-MS with a headspace extraction method was used to assess the volatile spectrum of microbiota metabolites. To assess the dynamic state of the microbiota, experiments were conducted on the formation of DSS-induced colitis in animals (rat/mice), followed by inflammation treatment by *Bacteroides fragilis* (JIM) vesicles, which are known to have a immunomodulatory properties.

Methods and Algorithms: For the formation of acute colitis and subsequent assessment of the spectrum of volatile metabolome at the stages of acute colitis and the stage of spontaneous convalescence, female Wistar rats from the Laboratory Animal Nursery (Pushchino RAS, Moscow region) at the age of 2 months were used. (200–230 g). Animals were randomly divided into three groups: control group (Control) (n = 5); experimental group, receiving dextran sodium sulfate for 7 days, followed by euthanasia (DSS7) (n = 5); group receiving dextran sodium sulfate followed by a week of recovery without DSS (DSS 14) (n = 5). Study of the therapeutic efficacy of *Bacteroides fragilis* (JIM) vesicles through comprehensive metabolomic and metagenomic analysis in an animal model of acute colitis was carried out on C57BL/6 mice, 7–8 weeks old. The mice were divided into three groups: k(n = 10), DSS (n = 10), DSS+V(n = 10). The experiment included two inflammation/remission cycles, each lasting 10 days. The DSS+V group received vesicles with a concentration of 1 mg/kg from days 11–20. Body weight, stool consistency, and presence of blood in stool were recorded every 3 days. Histological sample preparation was carried out according to the method described in publication [1]. Clinical manifestations of inflammation were assessed using the disease activity index (DAI). Metabolome analysis was carried out in stool samples using GC/MS on a Shimadzu QP2010 Ultra device with a Shimadzu HS-20 headspace extractor [2]. GC/MS data were processed using MetaboAnalyst 5.0 (http://www.metaboanalyst.ca) and GraphPad Prism 8.0.1 software. 16S rRNA gene sequencing for groups of experimental mice was performed on the MinION™ platform. Statistical comparison of groups was carried out using the non-parametric Mann–Whitney test and correction for multiple comparisons FDR. Calculations were carried out in the Python programming language, using the stats (scipy library) and stats.multitest (statsmodels library) modules. The plotting was carried out using the seaborn and matplotlib libraries.

Results: Based on the findings, the total DAI score included weight loss, stool consistency, and rectal bleeding. The DAI index in the control group of rats remained close to 0 throughout all experimental procedures. DAI in the groups of rats treated with DSS (DSS7 and DSS14) was significantly higher compared to the control group and increased on the 3rd day. The maximum score calculated in the groups of rats treated with DSS (DSS7 and DSS14) was obtained on day 7, when the acute phase of inflammation was reached. Animal weight loss, diarrhea and visible blood in the stool were observed on day 7 in both groups. According to histopathological examination, the colon of control animals had an intact mucosa, while in the colon of groups of DSS7 and DSS14 rats, inflammatory cell infiltration, thinning of the mucosa and submucosa and erosion of the epithelium were observed. As expected, DAI values decreased in the DSS14 group after stopping DSS administration on days 8–14. The ability of the intestines to recover was

clearly noted on the 14th day, when normalization of stool consistency and restoration of animal weight were observed. A certain improvement in the histological picture was observed in the DSS14 group, in which the number of crypts and goblet cells increased, and the structure of the mucous membrane was more differentiated, but minimal inflammatory infiltration remained. According to the GC/MS data obtained, changes in the metabolomic profile could be observed already on the third day after DSS administration. However, the main symptoms of colitis in rats, such as diarrhea and blood in the stool, have not yet been detected during this period of time. The acute phase of colitis in rats reached its peak on the 7th day after DSS administration and was characterized by changes in stool. Differences in the metabolomic profile were also recorded at day 7. The main changes affected compounds such as acetic acid, propanoic acid, propanoic acid-2-methyl-, butanoic acid, hexanoic acid. Despite improvement in histology after discontinuation of DSS administration, there was a trend toward progressive changes in the VOC profile. Metabolomic changes continued to increase by day 14. There were significant changes in the metabolomic profile for compounds such as acetic acid, butanoic acid, pentanoic acid, propanoic acid, 2-methyl-propanoic acid, hexanoic acid, reported in the DSS14 group. Changes in the relative amounts of metabolites such as benzenepropanoic acid and indole were noted at day 11.

In the second experiment, *Bacteroides fragilis* vesicles were previously isolated and characterized by TEM. The resulting vesicles were used to treat mice after exposure to DSS. All three groups underwent a histological examination of intestinal tissue on the 10th and 20th days of the experiment. According to histological examination, no changes were observed in the control group, whereas in the DSS group, the structure of the distal mucosa contained extensive foci of inflammatory lesions. In the DSS+V group, a smaller area of damage was observed, and the damage was focal in nature and had a significantly shorter duration. The DAI of the DSS and DSS+V groups was significantly higher than that of the control group, but during vesicle therapy, the DAI in the DSS+V group decreased significantly compared to the DSS group. The compounds identified in the metabolomics study included short-, medium- and long-chain fatty acids and amino acid derivatives. On the 10th day of the experiment, no significant differences were found between the DSS and DSS+V groups. At day 20, corresponding to the end of vesicle treatment, three distinct groups could be distinguished. The control group differed significantly from the DSS and DSS+V groups in terms of the metabolomic spectrum. When treated with vesicles in the DSS+V group, a tendency could be observed to restore the overall metabolite profile to control values. Analysis of microbiome variability in stool samples on day 10 of the experiment demonstrated a decrease in the abundance of certain bacterial species compared to the control group. By the 20th day, a significant difference was observed between the DSS groups and the control group, while the relative abundance of the studied bacterial genera in the DSS+V group reached the initial values determined in the control group by the twentieth day of observation. Correlation analysis revealed that the increase in *Saccharibacteria* and *Acetivibrio* was positively correlated with phenol and pentanoic acid, and in the DSS group, the predominance of *Lactococcus* and *Romboutsia* was negatively correlated with phenol and pentanoic acid.

Conclusions: model experiments show that the symbiosis of two techniques (GC/MS and metagenomic sequencing) allows non-invasive monitoring of all stages of inflammation and restoration processes of the intestinal epithelium in experimental animals. Thus, functional metabolomics data can be used to monitor inflammatory activity in patients with IBD.

Список литературы/References

1. Shagaleeva O.Yu., Kashatnikova D.A., Vorobyeva E.A. et al. Therapeutic effects of *Bacteroides fragilis* vesicles in a model of chemically induced colitis in rats. *Bull Exp Biol Med*. 2024;177(5):586-590. doi 10.47056/0365-9615-2024-177-5-586-590
2. Shagaleeva O.Y., Kashatnikova D.A., Kardonsky D.A. et al. GC-MS with headspace extraction for non-invasive diagnostics of IBD dynamics in a model of DSS-induced colitis in rats. *Int J Mol Sci*. 2024;25(6):3295. doi 10.3390/ijms25063295

doi 10.18699/bgrs2024-8.1-12

Application of machine learning algorithms to study the gut microbiota of patients with depressive disorders

Galanova O.O.[1, 2]*, Kovtun A.S.[1]

[1] *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, State University, Dolgoprudny, Russia*

* *galanova.oo@phystech.edu*

Key words: depression; gut-microbiota; gut-brain axis; machine learning

Motivation and Aim: Depression remains one of the most important health problems of our time due to its widespread and devastating impact on the human psyche. According to the World Health Organization, 5 % of adults worldwide suffer from depression. Scientific research usually looks at the biochemical and neural aspects of a given disease. At the same time, the potential impact of the gut microbiota (GM) is often not taken into account, despite its enormous metabolic potential. Gut bacteria are able to influence the development and functioning of various human organs and systems, including the nervous system through two-way communication implemented through the gut-brain axis [1]. It means that disorders in this complex system may have a significant impact on human mental health, and vice versa [2, 3]. Studies of GM metabolites with neuromodulating activity affecting mental health are becoming increasingly relevant [4, 5]. In this context, understanding of the role of GM opens up the possibility of developing more effective methods for the diagnosis and treatment of depression, as well as other mental disorders [1, 4, 6]. Identifying and analyzing the relationships between the development of mental illness in humans and changes in GM, both taxonomic and functional, is a complex task that requires processing and interpretation huge amounts of data. Machine learning provides a promising approach to analyzing large-scale metagenomic data and identifying taxonomic and functional biomarkers associated with various diseases, including depressive disorders. The general idea of many ML algorithms is to train a model based on a list of parameters and a special training dataset and then use it for the analysis of a general dataset. One of the most important parts of application of the ML is the assembly of an appropriate training dataset, which avoids such problems as biases, batch effect, noisiness and dataset size. Random forest classification is one of the most widely used algorithms in metagenomics thanks to its ability to work with relatively small datasets while delivering stable results. With the help of various ML approaches, researchers can solve the tasks of taxonomic characterization, sample clusterization and correlation of microbiota composition with various diseases. A number of studies have shown a correlation between changes in GM and major depressive disorder [7–9]. The results obtained using machine learning may contribute to a better understanding of the role of GM in major depressive disorder and pave the way for new diagnostic and therapeutic strategies.

Methods and Algorithms: A cohort of 74 samples of gut microbiota, obtained from 36 patients with depression and 38 healthy controls from Moscow Region, was used for the research. The metagenomic analysis was carried out using the signature approach that allows for identification of both genes, their bacterial source and abundance of the pairs (gene; taxon). For the application of the machine learning approach, the cohort was first divided into two groups, training and validation, and then broadened with the

simulated data generated using MOSTLY AI tool based on GAN. The models were generated using several machine learning algorithms, including random forest, elastic net, and YOLO.

Results: Compositional changes in the GM of patients with MDD and subsequent changes in its neurometabolic potential were indicated. The results demonstrated a decrease in beneficial bacteria, such as *Faecalibacterium*, *Roseburia*, and *Lachnospira*, along with an increase in abundances of opportunistic pathogens, such as *Ruthenibacterium* and *Escherichia* in the PwD group compared with healthy controls. These changes characterized the dysbiosis of GM and could be contributing factors to the development of depression. The signature approach at the "Species" level revealed a set of genes with significantly decreased abundance in the patients' group, most of which were identified in species *Faecalibacterium prausntizii*. These data represent the neurometabolic signature of the GM of depressive patients and can be used as biomarkers. Further application of the machine learning algorithms allowed for creation of the models that can determine the depressed condition of a patient using the information about their metagenomics signature. The best results were observed for the models generated by the random forest algorithm with the prediction accuracy of 80 % and YOLO with the prediction accuracy of more than 90 %.

Conclusion: Application of machine learning algorithms in metagenomics studies is gaining popularity. They can be used for classification tasks introducing novel approaches for taxonomic annotation of microbiota, and in association studies for identification of correlations between changes in microbiota and the host's condition, including such disorders, as depression. In our research we have made a complex description of changes of gut microbiota, which can be connected with the depressed condition. The obtained models indicate that the specific genes and taxa of the gut microbiota can be viewed as the novel group of biomarkers for diagnostics of depression in addition to the currently used methods. The currently developed models can already be used for the *in silico* diagnostics using the metagenomics sequencing and the signature approaches. The development of new methods of treatment for the depression by the correction of gut microbiota with the use of psychobiotics is another possible application for these algorithms.

Funding: The study is supported by the Russian Science Foundation (№ 20-14-00132).

References

1. The neuroscience of depression genetics, cell biology, neurology, behavior, and diet. Eds. Martin C.R., Hunter L.-A. London, UK, 2021
2. Sasso J.M. et al. Gut microbiome-brain alliance: a landscape view into mental and gastrointestinal health and disorders. *ACS Chem Neurosci.* 2023;14:1717-1763. doi 10.1021/acschemneuro.3c00127
3. Kovtun O., Leneva N., Bykov Y.S. et al. Structure of the membrane-assembled retromer coat determined by cryo-electron tomography. *Nature*. 2018;561:561-564. doi 10.1038/s41586-018-0526-z
4. Poluektova E.U., Yunes R.A., Danilenko V.N. the putative antidepressant mechanisms of probiotic bacteria: relevant genes and proteins. *Nutrients*. 2021;13:1591. doi 10.3390/nu13051591
5. Averina O.V., Zorkina Y.A., Yunes R.A. et al. Bacterial metabolites of human gut microbiota correlating with depression. *Int J Mol Sci*. 2020;21:9234. doi 10.3390/ijms21239234
6. Liu L., Wang H., Chen X., Zhang Y., Zhang H., Xie P. Gut microbiota and its metabolites in depression: From pathogenesis to treatment. *EBioMedicine*. 2023;90:104527. doi 10.1016/j.ebiom.2023.104527
7. Angelova I.Y. et al. Unveiling the connection between microbiota and depressive disorder through machine learning. *Int J Mol Sci*. 2023;24(22):16459
8. Radjabzadehc D. et al. Gut microbiome-wide association study of depressive symptoms. *Nat Comm.* 2022;13:7128. doi s41467-022-34502-3
9. Liang S. et al. Multi-cohort analysis of depression-associated gut bacteria sheds insight on bacterial biomarkers across populations. *Cell Mol Life Sci*. 2022;80(1):9. doi 10.1007/s00018-022-04650-2

doi 10.18699/bgrs2024-8.1-13

Honey is a natural metatbiotic (postbiotic) product

Ilyasov R.A.[1]*, Danilenko V.N.[2], Ilyasova A.Y.[1], Boguslavsky D.V.[1]

[1] *Koltsov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, Russia*

[2] *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia*

* *apismell@hotmail.com*

Key words: microbiome; honey bee; lactobacteria; Apis mellifera; Lactobacillus; postbiotics; metabiotics

The antioxidant effects of honey can be primarily attributed to its phenolic constituents, which have been demonstrated to provide protection to human cells and the bloodstream when consumed [1]. Analogous to its antibacterial activity, the antioxidant capacity of honey varies considerably and is dependent on the floral source from which it is derived. Darker-colored honey has been shown to possess greater antioxidant activity compared to lighter-colored honey, as the color is influenced by the phenolic content [1]. Furthermore, the phenolic components of honey are also closely connected to its anti-inflammatory qualities. Extant research has revealed that honey has the ability to interrupt inflammatory mediators and decrease pro-inflammatory cytokine levels [1]. Consequently, the anti-inflammatory and antioxidant activities of honey appear to be intrinsically linked [2]. Certain honeys are frequently employed in the treatment of wounds, burns, and ulcers due to their documented anti-inflammatory, antioxidant, antibacterial, and wound-healing properties [3]. However, the systemic effects of honey when consumed orally remain relatively understudied. Here we aimed to investigate the beneficial properties of honey as a natural metabiotic (postbiotic), including its microbe metabolites, plant waste products, and the contributions of bees.

Honey, the primary product of beekeeping, has been a part of human consumption since ancient times [4]. In general, plant nectar is necessary for the production of honey. The glands in plant flowers create nectar, a tasty liquid fluid [5]. Through enzymatic characteristics, the productive microbiota of honey bees and their own enzyme, invertase, affect the conversion of nectar into honey [5]. The fermentation process turns honey into a postbiotic substance [6]. Honey bees stimulate the enzyme glucose oxidase, which converts glucose into gluconic acid and hydrogen peroxide, as honey replaces the nectar [3]. These elements play a crucial role in determining the taste and biological functions of honey [3]. Depending on the specific process used to convert the nectar, the presence and concentration of acids and other compounds may vary, resulting in some honeys exhibiting stronger antibacterial properties than others [3]. The co-fermentation of nectar and nectar modification by bee microbiota and stomach enzymes result in the production of honey [5]. In addition to the active ingredients, phytohormones, and polyphenols originating from plant nectar, such as phenolic acids, flavonoids, and other phytochemicals, honey also contains a number of compounds created by fermentation in the ventricle [7]. These compounds can influence antioxidant activity, neuromodulation, and inflammation [8]. Metabiotics (postbiotics) are defined as the compounds released by bacteria as a result of their metabolic activities, and they may also include dead bacteria that can be beneficial to the human body [6]. Considering these characteristics, honey should be recognized as a natural metabiotic (postbiotic) substance that can be utilized by humans to treat a variety of health conditions [4].

Funding: The study is supported by IDB RAS Government basic research program in 2024 No. 0088-2024-0009.

References

1. Schell K.R. et al. The potential of honey as a prebiotic food to re-engineer the gut microbiome toward a healthy state. *Fronti Nutr.* 2022;9:957932. doi 10.3389/fnut.2022.957932
2. Danilenko V.N., Devyatkin A.V., Marsova M.V., Shibilova M.U., Ilyasov R.A., Shmyrev V.I. Common inflammatory mechanisms in COVID-19 and Parkinson's diseases: the role of microbiome, pharmabiotics and postbiotics in their prevention. *J Inflamm Res.* 2021;14:6349-6381. doi 10.2147/JIR.S333887
3. Molan P. Why honey is effective as a medicine. *Bee World.* 2015;82:22-40. doi 10.1080/0005772x.2001.11099498
4. Ilyasov R.A. et al. Chapter 3. Honey: a natural postbiotic product. In: Stillman E.L. (Ed.). Exploring microbiotics in health and disease. Postbiotics, Probiotics, and Prebiotics. Nova Science Publishers, 2024
5. Schramm D.D. et al. Honey with high levels of antioxidants can provide protection to healthy human subjects. *J Agricult Food Chemistry.* 2003;51:1732-1735. doi 10.1021/jf025928k
6. Zhao H. et al. Honey polyphenols ameliorate dss-induced ulcerative colitis via modulating gut microbiota in rats. *Mol Nutr Food Res.* 2019;63:e1900638. doi 10.1002/mnfr.201900638
7. Vallianou N.G. Honey and its anti-inflammatory, anti-bacterial and anti-oxidant properties. *Genl Med.* 2014;2(2):1000132. doi 10.4172/2327-5146.1000132

8 Симпозиум «Микробиология и биотехнологии: компьютерные и экспериментальные подходы»

Symposium “Microbiology and biotechnologies: computational and experimental approaches”

doi 10.18699/bgrs2024-8.2-01

Изучение физиологических групп микроорганизмов метанокисляющих сообществ

Бабусенко Е.С., Лалова М.В.*, Нюньков П.А., Цымбал В.В.
ООО «Гипробиосинтез», Москва, Россия
** lalova.m@gibios.ru*

Ключевые слова: сообщества метанокисляющих бактерий; экологические стратегии; мультисубстратное тестирование

Метанокисляющие бактерии (метанотрофы) – уникальная группа бактерий, отличающаяся способностью использовать метан в качестве единственного источника углерода и энергии. Эти бактерии рассматриваются как перспективный объект для экологически чистого производства микробной биомассы на природном газе с целью получения кормовых добавок, для выделения из биомассы биологически активных соединений.

В природной среде метанотрофы склонны образовывать сложные ассоциации с другими бактериями. В состав таких ассоциаций, помимо, метанокисляющих бактерий, входят одна или более синтрофных культур, потребляющих либо продукты метаболизма данных бактерий, либо продукты, образующиеся в результате работы основного фермента метанмонооксигеназы. Такие сообщества позволяют более полно ассимилировать газообразные субстраты, находящиеся в природном газе, что предотвращает ингибирование ими основного продуцента и, следовательно, обеспечивает стабильность и эффективность всего сообщества.

Однако в условиях незащищенной ферментации самопроизвольно формируются смешанные культуры, состоящие из основного производственного штамма и сопутствующей микрофлоры.

Изучение состава природных сообществ, их трофических связей, характера взаимоотношений между видами в биоценозе, является основой для разработки технологических приемов и способов сохранения устойчивости смешанных культур.

Из объектов окружающей среды уже выделено значительное количество метанокисляющих бактерий, однако число высокопродуктивных культур, использование которых экономически целесообразно, невелико.

Цель: поиск новых метанокисляющих бактерий и их сообществ, способных потреблять метан из природного газа с удовлетворительным выходом биомассы.

Были проведены работы по выделению метанокисляющих бактерий из природных образцов, отобранных на Южно-Кисловском месторождении. Использовали конденсат газовый стабильный (КГС), газовый конденсат нестабильный (КГН), почву, отобранную на площадке, расположенной на удалении 0.5 км от газоносной скважины (ПГС).

Выделение метанокисляющих бактерий проводили, используя метод накопительных культур.

Для получения накопительных культур использовали минеральную среду «П» [1]. Образцы проб были отобраны с глубины скважины газового месторождения, имеющей температурный фон 55 °С.

Выделение метанотрофов проводили при температуре: 55, 50, 45, 40, 30 °С. В колбы объемом 1000 мл вносили 200 мл минеральной среды и 30 мл природного образца. Засеянные колбы параллельно заполняли либо чистым метаном (99.9 %), либо природным газом (содержание метана не менее 96 %), затем инкубировали на качалке при 220 об/мин.

После первых трех пассажах при температурах 30, 40 и 45 °С в образцах КГС и КГН наблюдалось незначительное помутнение среды при росте на чистом метане. Но при последующих пересевах роста не было. На природном газе признаки роста отсутствовали при инкубировании при всех указанных температурах.

При использовании природного газа в образце ПГС был отмечен стабильный рост при температурах 30 и 45 °С, причем на природном газе рост был более интенсивный, чем на чистом метане. Для каждой температуры было проведено 12 пассажей по 5 суток каждый.

На 5–7-е сутки в колбах появлялся заметный диффузный или хлопьевидный рост. Как правило, метанокисляющие культуры являют собой ассоциацию метанотрофных и сопутствующих микроорганизмов. В первичных накопительных культурах может присутствовать несколько (2–3) видов метанотрофов и до 7–8 бактериальных спутников, окисляющих метанол и другие продукты соокисления гомологов метана, а также использующих продукты метаболизма метанотрофа [1].

При микроскопировании суспензии (I), полученной при 30 °С, наблюдали значительное разнообразие бактериальной микрофлоры. Наряду с преобладанием кокков и диплококков присутствовали палочковидные бактерии различных размеров. В суспензии, полученной при 40 °С (II), микробное разнообразии было меньше.

Структура ассоциации микроорганизмов и ее физиологические параметры изменяются во времени и пространстве, подвержены значительному влиянию различных факторов окружающей среды.

Для понимания структуры микробного сообщества, характера взаимодействия отдельных видов микроорганизмов, входящих в состав сообщества необходимо исследовать таксономическое и функциональное разнообразие данной ассоциации микроорганизмов.

Традиционные методы применения элективных питательных сред позволяют учитывать численность микроорганизмов лишь на первых этапах сукцессии, когда велико разнообразие микроорганизмов в системе. Кроме того, таксономический состав микроорганизмов, представляющих ту или иную группу микроорганизмов, непостоянен и изменяется в зависимости от конкретных условий. Поэтому физиологические группы, выявляемые на определенных элективных средах, совсем не обязательно осуществляют те же процессы в производственных условиях.

Существуют более прогрессивные методы, например, метод анализа спектра потребления органических субстратов (СПС) природной ассоциацией [2]. Разработаны модификации этого метода [3], которые позволяют определить не только спектр используемых микроорганизмами субстратов, но и дать количественную оценку потребления каждого из них.

Устойчивость тех или иных физиолого-биохимических функций ассоциативных микроорганизмов при изменении физико-химических параметров обусловлена наличием нескольких дублирующих друг друга видов микроорганизмов, с одной стороны, и их адаптацией к новым условиям – с другой [4].

Для характеристики функционального разнообразия микробного сообщества был апробирован метод мультисубстратного тестирования (МСТ), который был выполнен в соответствии со стандартной методикой [3]. При проведении анализа оценивали спектры потребления субстратов микробными сообществами I и II. В работе использовали стандартные 48-луночные планшеты с 11 источниками органического углерода (альдегиды, спирты, органические кислоты, углеводы). Две ячейки были контрольными. Данные планшеты с субстратами были приготовлены на кафедре биологии почв факультета почвоведения МГУ им. М.В. Ломоносова. Планшеты инкубировали в термостате в течение 3-х суток при температуре 28 °С до появления визуально регистрируемой окраски ячеек. При росте микроорганизмов, потребляющих субстрат, в ячейках происходило восстановление неокрашенных солей тетразолия до бордово-окрашенного формазана. По завершении инкубации проводили фотометрическое считывание оптической плотности ячеек при длине волны 510 нм.

При обработке результатов МСТ оценивали количество используемых субстратов и интенсивность их потребления. Установлено, что существуют количественные и качественные различия в ассимиляции субстратов. Так, микробное сообщество I (выделенное при температуре 30 °С) ассимилировало 9 субстратов (метанол, этанол, бутанол, ацетат, формальдегид, глюкозу, галактозу, арабинозу, фруктозу), а микробное сообщество II (выделенное при 45 °С) – 4 субстрата (этанол, формальдегид, ацетат, глюкозу).

Таким образом, получены два микробных метанокисляющих сообщества, которые можно рассматривать, как перспективные объекты для промышленного получения кормовой добавки на основе метанокисляющих бактерий. Следующий этап работы предполагает выделение из полученных сообществ чистых культур, их идентификацию и изучение физиолого-биохимических свойств.

Список литературы/References

1. Гальченко В.Ф. Метанотрофные бактерии. М.: ГЕОС, 2001
2. Garland J.L. Analysis and interpretation of community-level physiological profiles in microbial ecology. *FEMS Microbiol Ecol.* 1997;24:289-300. doi 10.1111/j.1574-6941. 1997.tb00446.x
3. Горленко М.В., Кожевин П.А. Мультисубстратное тестирование природных микробных сообществ. М.: МАКС Пресс, 2005
4. Звягинцев Д.Г. Почва и микроорганизмы. М.: Изд-во Моск. ун-та, 1987

doi 10.18699/bgrs2024-8.2-02

Геномика и физиология нитчатых бесцветных серобактерий

Грабович М.Ю.[1]*, Руденко Т.С.[1], Смольяков Д.Д.[1], Равин Н.В.[2]
[1] *Воронежский государственный университет, Воронеж, Россия*
[2] *Институт биоинженерии, ФИЦ Биотехнологии РАН, Москва, Россия*
* *margarita_grabov@mail.ru*

Ключевые слова: бесцветные серобактерии; *Thiothrix*; *Beggiatoa*; метагеном; таксономия

Мотивация и цель: Бесцветные серобактерии – это одна из загадочных групп литотрофных прокариот. Многие из них характеризуются гигантскими размерами (20–750 мкм), среди них преобладают некультивируемые формы, они откладывают элементную серу внутри клеток (до 80 % от сухого веса), в природных биотопах, как правило, доминируют, образуя мощные видимые невооруженным глазом маты и обрастания [1].
Представительная коллекция изолятов рода *Thiothrix,* полученная нашей группой, позволила выявить проблему в идентификации изолятов: большинство культур имели высокий уровень гомологии по 16S rRNA, которая варьирует в пределах 98–99 %, что позволяет предположить, что они являются штаммами одного вида. Эта проблема усложнялась тем, что представителей рода *Thiothrix* объединяет схожий фенотип – нитчатые бесцветные серобактерии, образуют розетки и внутриклеточные включения элементной серы при росте в присутствии сероводорода и тиосульфата, и, как правило, все представители рода имеют унифицированную схему основных метаболических путей [2]. Проблема частично решилась с появлением возможности сборки геномов (MAG) из метагеномов. Однако иногда наблюдаемое присутствие в образце нескольких близких, но генетически отличающихся штаммов, препятствует как сборке метагеномных контигов, так и их корректному биннингу для сборки MAG. Таким образом, отсутствие достоверных филогенетических маркеров и небольшое количество изолятов препятствуют установлению точного филогенетического родства в роде *Thiothrix* [3, 4].
За последнее десятилетие семейство *Beggiatoaceae* было значительно расширено во многом благодаря данным метагеномного анализа. Однако описание новых таксонов в семействе преимущественно основано на морфологии, а геномные данные не содержат эту информацию. Это, в свою очередь, привело к возникновению большого количества новых кандидатных родов, а многие из уже описанных отсутствуют в современной геномной таксономии этого семейства (Genome Taxonomny Database, GTDB).
Методы и алгоритмы: Метагеномное секвенирование и сборка геномов, геномный анализ и аннотация, филогенетический анализ.
Результаты: Мы проанализировали возможность использования различных консервативных генов (около 100) в качестве филогенетических маркеров для рода *Thiothrix,* отличных от 16S rRNA. Установлено, что уровни сходства нуклеотидных последовательностей генов тРНК(Ile)-лизидинсинтазы (*tilS*) и β-субъединицы РНК-полимеразы (*rpoB*) хорошо согласуются со средними

значениями нуклеотидной идентичности (ANI) между геномами различных представителей рода *Thiothrix*. Топологии филогенетических деревьев, построенных на основе последовательностей *tilS* и *rpoB*, сходны с топологией дерева, построенного на основе последовательностей 120 консервативных маркерных генов. Высокопроизводительное секвенирование фрагментов гена *tilS* было использовано для профилирования микробного мата с *Thiothrix* из сульфидного источника, что позволило выявить известные виды *Thiothrix* и новых филотипы видового уровня. Таким образом, использование *tilS* и *rpoB* в качестве филогенетических маркеров позволит проводить экспресс-анализ чистых культур и природных сообществ с целью филогенетической идентификации представителей рода *Thiothrix*.

Создана полногеномная филогения рода *Thiothrix*, основанная на широкой выборке секвенированных геномов. К началу наших исследований филогения рода *Thiothrix* была в зачаточном виде, так как видовое разнообразие было представлено 5 видами, а геномы были просеквенированы лишь для 3 видов. С применением филогенетического анализа, основанного на полногеномных последовательностях, нам удалось расширить род до 15 видов, для 9 из которых нами определены геномные последовательности. Также нами был описан новый род в составе семейства *Thiothrichaceae* '*Candidatus* Thiocaldithrix' gen. nov.

Проведен пангеномный анализ рода *Thiothrix*. Коровый геном включает гены диссимиляционного метаболизма серы и центральных метаболических путей; вспомогательный геном – гены диссимиляционной нитрат- и тиосульфат-редукции, фиксации азота; уникальные гены – гипотетические белки, транспортеры, регуляторы транскрипции, метилтрансферазы, транспозазы и системы токсин-антитоксин. Впервые показана способность большинства представителей рода *Thiothrix* к анаэробному дыханию в присутствии нитратов и тиосульфата, а также наличие у некоторых представителей рода *Thiothrix* «натриевой» биоэнергетики, включая Na^+-АТРазу. Можно предполагать, что дуалистичность энергетики позволяет этим бактериям в микроаэробных и анаэробных условиях перейти на более экономичный энергорежим за счет того, что биомембраны имеют меньшую утечку по натрию, чем по протону, и, следовательно, дольше сохраняют электрохимический градиент ионов Na^+ на мембране в микроаэробных и анаэробных условиях.

Семейство *Beggiatoaceae* в настоящее время представлено 25 родами в GTDB, из которых только шесть имеют определенный таксономический статус. В результате метагеномного анализа бактериальных матов, скапливающихся в ламинариевой «Помойке» на глубине 15–18 м в Белом море, удалось получить два MAGs. На основе проведенного филогенетического анализа один из них, MAG WS_Bin3, был описан как новый род в семействе *Beggiatoaceae* с предложенным названием '*Candidatus* Albibeggiatoa psychrophila' gen., nov., sp. nov WS_Bin3. Таким образом, идентифицирован кандидатный род «BB20», в составе которого находится WS_Bin3 с двумя дополнительными видами.

Второй MAG, WS_Bin1, оказался штаммом ранее описанного вида '*Ca*. Parabeggiatoa communis' в составе ранее описанного рода '*Ca*. Parabeggiatoa' [5]. Учитывая принадлежность MAG WS_Bin1 к роду '*Ca*. Parabeggiatoa' на основе гомологии гена 16S рРНК, кандидатный род «UBA10656» был определен как '*Ca*. Parabeggiatoa' с тремя новыми видами помимо '*Ca*. Parabeggiatoa communis'. Сравнительный анализ геномов позволил выявить основные сходства и различия

в метаболизме у представителей родов ‘*Candidatus* Parabeggiatoa’ и ‘*Candidatus* Albibeggiatoa’.
Выводы: Обобщены многолетние исследования по изучению генетических основ природного разнообразия, физиологии и таксономии одной из групп бесцветных серобактерий рода *Thiothrix*. Для создания достоверной филогении рода были определены геномы 9 видов рода *Thiothrix*. С учетом известных и новых геномных последовательностей была предложена обновленная филогения рода *Thiothrix*. Филогения, основанная на использовании полногеномных последовательностей, позволила внести ясность в таксономию рода *Thiothrix*. На основе полногеномного анализа создана обновленная филогения для семейства *Beggiatoaceae*, в частности для родов ‘*Candidatus* Parabeggiatoa’ и ‘*Candidatus* Albibeggiatoa’, а также внесена ясность в понимание метаболического потенциала бесцветных серобактерий из этих родов.
Получение чистых культур и их геномных сиквенсов помогли бы совершить прорыв в таксономии бесцветных серобактерий, но подавляющее большинство бактерий этой группы относятся к группе сложно культивируемых, а многие из них являются некультивируемыми формами. Поэтому на данном этапе лишь метагеномный подход наряду с полевыми наблюдениями и микроскопическим анализом морфологии бактерий будут способствовать разрешению таксономии бесцветных серобактерий.
Финансирование: Исследование поддержано Российским научным фондом (№ 20-14-00137, https://rscf.ru/en/project/20-14-00137/).

Genomics and physiology of filamentous colorless sulfur bacteria

Grabovich M.Yu.[1]*, Rudenko T.S.[1], Smolyakov D.D.[1], Ravin N.V.[2]
[1] *Voronezh State University, Voronezh, Russia*
[2] *Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russia*
* *margarita_grabov@mail.ru*

Key words: colorless sulfur bacteria; *Thiothrix*; *Beggiatoa*; metagenome; taxonomy

Motivation and Aim: Colorless sulfur bacteria are one of the most interesting groups of lithotrophic prokaryotes. Many of them are characterized by giant sizes (20–750 μm), uncultured forms prevail among them, it deposit elemental sulfur intracellularly (up to 80 % of their dry weight), and in natural biotopes, as a rule, dominate, forming thick mats and fouling visible to the naked eye [1].
A representative collection of isolates of the genus *Thiothrix* obtained by our group revealed a problem in the identification of isolates: most bacterial cultures had a high level of 16S rRNA homology, which varies between 98–99 %, suggesting that they are strains of the same species. This problem is complicated by the fact that representatives of the genus *Thiothrix* share a similar phenotype: filamentous colorless sulfur bacteria forming rosettes and intracellular inclusions of elemental sulfur under growth in the presence of hydrogen sulfide and thiosulfate, and as a rule, all representatives of the

genus have a unified scheme of basic metabolic pathways [2]. This problem was partially solved with the advent of the ability to assemble MAGs from metagenomic sequences. However, the frequently observed presence of several similar but genetically different strains in an environmental sample prevents both the assembly of metagenomic contigs and their correct binning into MAGs. Thus, the lack of reliable phylogenetic markers made it difficult to determine the exact phylogenetic relationship of the already described species, as well as that of new isolates [3, 4].

The family *Beggiatoaceae* has greatly expanded over the last decade, mainly due to metagenomic analysis. However, the description of new taxa in the family is predominantly based on morphology, while genomic data, in contrast, are devoid of this information. This, in turn, has led to a large number of new unidentified candidate genera, while many of the already described genera are absent from the current taxonomy of the family according to the Genomic Taxonomy Database (GTDB).

Methods and Algorithms: Metagenome sequencing and assembly of MAGs, genome analysis and annotation, phylogenetic analysis.

Results: We analyzed the possibility of using various conserved genes (about 100) as phylogenetic markers for the genus *Thiothrix* different from 16S rRNA. We found that the nucleotide sequence similarity levels of tRNA(Ile)-lysidine synthase (*tilS*) and β-subunit RNA polymerase (*rpoB*) genes align well with the average nucleotide identity (ANI) values between the genomes of different representatives of the genus *Thiothrix*. The topologies of phylogenetic trees based on *tilS* and *rpoB* sequences are similar to those of a tree based on the sequences of 120 conserved marker genes. High-throughput sequencing of *tilS* gene fragments was used to profile microbial mat with *Thiothrix* from a sulfide source, which made it possible to identify known *Thiothrix* species and new species-level phylotypes. Thus, the use of *tilS* and *rpoB* as phylogenetic markers will allow express analysis of pure cultures and natural communities for phylogenetic identification of representatives of the genus *Thiothrix*.

We have constructed a full-genome phylogeny of the genus *Thiothrix* based on a wide sample of sequenced genomes. At the beginning of our studies, the phylogeny of the genus *Thiothrix* was in its infancy, as species diversity was represented by 5 species, and the genomes have been sequenced for only 3 species. Using phylogenetic analysis based on full-genome sequences, we were able to expand the genus to 15 species, for 9 of which we have identified genomic sequences. species. We also described a new genus in the family *Thiothrichaceae* '*Candidatus* Thiocaldithrix' gen. nov. that is phylogenetically close to the genus *Thiothrix*.

We obtained a pangenome of the genus *Thiothrix*. The core genome includes genes of dissimilatory sulfur metabolism and central metabolic pathways; the auxiliary genome includes genes of dissimilatory nitrate and thiosulfate reduction, nitrogen fixation; unique genes include hypothetical proteins, transporters, transcription regulators, methyltransferases, transposases, and toxin-antitoxin systems.

We have shown for the first time the ability of most representatives of the genus *Thiothrix* to anaerobic respiration in the presence of nitrate and thiosulfate, as well as the presence of "sodium" bioenergetics, including Na+- ATPase, in some representatives of the genus *Thiothrix*. It can be assumed that the duality of energy allows these bacteria, under microaerobic and anaerobic conditions, to transition to a more economical energy regime to a more economical energy regime due to the fact that biomembranes have less leakage by sodium than by proton and consequently keep the electrochemical gradient of Na^+ ions on the membrane longer.

The family *Beggiatoaceae* is currently represented by 25 genera in the GTDB, of which only six have a definite taxonomic status. Two MAGs were assembled as a result of metagenome analysis of bacterial sulfur mats formed on decaying laminaria remnants in the sea bay "Laminaria dump" at a depth of 15–18 m in the White Sea. Based on phylogenetic analysis, one of them, MAG WS_Bin3, was assigned to a new genus and species in the family *Beggiatoaceae* and named '*Candidatus* Albibeggiatoa psychrophila' gen., nov., sp. nov. Accordingly, the candidate genus 'BB20' was identified, which includes three species.

The other MAG, namely WS_Bin1, turned out to be a strain of the previously described species '*Candidatus* Parabeggiatoa communis' within the earlier described genus '*Candidatus* Parabeggiatoa' [5]. Considering that MAG WS_Bin1 belongs to the genus '*Ca*. Parabeggiatoa' on the basis of 16S rRNA gene homology, the candidate genus 'UBA10656' was assigned to the genus '*Ca*. Parabeggiatoa', comprising three new species in addition to the species '*Ca*. Parabeggiatoa communis'.

Comparative analysis of genomes revealed the main similarities and differences in the metabolism of the genera '*Ca.* Parabeggiatoa' and '*Ca.* Albibeggiatoa'.

Conclusion: Long-term studies on the genetic basis of natural diversity, physiology and taxonomy of one of the groups of colorless sulfur bacteria of the genus *Thiothrix* are summarized. To create a reliable phylogeny of the genus, the genomes of 9 species of the genus *Thiothrix* were determined. Based on the known and new genomic sequences, an updated phylogeny of the genus *Thiothrix* was proposed. A phylogeny based on the use of full-genome sequences has provided clarity to the taxonomy of the genus *Thiothrix*. An updated phylogeny for the family *Beggiatoaceae* was established based on a full-genome analysis, in particular, for the genera '*Ca.* Parabeggiatoa' and '*Ca.* Albibeggiatoa', was developed. In addition, the insight into the metabolic potential of the colorless sulfur bacteria of these genera was clarified.

Obtaining pure cultures and their genome sequences could help to make a breakthrough in the taxonomy of colorless sulfur bacteria. However, an obstacle to this task is the difficulty of culturing or even the impossibility of culturing the vast majority of bacteria of this group. Therefore, at present, only a metagenomic approach along with field observations and microscopy analysis of bacterial morphology will help to solve the problem of taxonomy of colorless sulfur bacteria.

Funding: The study is supported by the Russian Science Foundation (No. 20-14-00137, https://rscf.ru/en/project/20-14-00137/).

Список литературы/References

1. Nelson D.C., Castenholz R.W. Use of reduced sulfur compounds by *Beggiatoa* sp. *J Bacteriol*. 1981;147(1):140-154. doi 10.1128/jb.147.1.140-154.1981
2. Grabovich M.Y., Ravin N.V., Boden R. Thiothrix. In: Bergey's Manual of Systematics of Archaea and Bacteria. John Wiley & Sons, 2023; 21
3. Howarth R., Unz R.F., Seviour E.M. et al. Phylogenetic relationships of filamentous sulfur bacteria (*Thiothrix* spp. and Eikelboom type 021N bacteria) isolated from wastewater-treatment plants and description of *Thiothrix eikelboomii* sp. nov., *Thiothrix unzii* sp. nov., *Thiothrix fructosivorans* sp. nov. and *Thiothrix defluvii* sp. nov. *Int J Syst Bacteriol*. 1999;49(4):1817-1827. doi 10.1099/00207713-49-4-1817
4. Rossetti S., Blackall L.L., Levantesi C., Uccelletti D., Tandoi V. Phylogenetic and physiological characterization of a heterotrophic, chemolithoautotrophic *Thiothrix* strain isolated from activated sludge. *Int J Syst Evol Microbiol*. 2003;53(5):1271-1276. doi 10.1099/ijs.0.02647-0
5. Salman V., Amann R., Girnth A.C. et al. A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. *Syst Appl Microbiol*. 2011;34(4):243-259. doi 10.1016/j.syapm.2011.02.001

doi 10.18699/bgrs2024-8.2-03

Пайплайн для обработки данных метабаркодинга грибов природных сообществ

Ишманов Т.[1]*, Звягина Е.[1, 2], Рудыкина Е.[1], Филиппов И.[1], Бульонкова Т.[1], Добрынина А.[1], Филиппова Н.[1]

[1] *Югорский государственный университет, Ханты-Мансийск, Россия*

[2] *Московский государственный университет, Москва, Россия*

** ishmanov2003tagir@mail.ru*

Ключевые слова: метабаркодинг; eDNA; микобиота; DNA-derived-data; Мухрино; Западная Сибирь

Мотивация и цель: Оценка разнообразия микробиомов природных сообществ в последнее десятилетие претерпела существенный рывок с развитием методов метабаркодинга. На данный момент для обработки данных секвенирования eDNA критически необходима оптимизация набора и параметров работы отдельных модулей обработки под особенности полученных данных (наличие остатков адаптеров, праймеров и т. д.) и систематическую группу исследуемых организмов (растения, грибы, беспозвоночные, …).

Для использования метабаркодинга в анализе структуры разнообразия грибных сообществ природных экосистем был собран и оптимизирован пайплайн, позволяющий эффективно обработать последовательности ITS2 грибов, полученные на платформе Illumina MiSeq.

Методы и алгоритмы: Пайплайн подготовлен на Python 3 в интерфейсе Artifact API для QIIME 2 [1], который позволяет модифицировать параметры в зависимости от целей исследования и качества сырых данных. Этапы пайплайна: 1) подача файлов fastq; 2) поиск и удаление адаптеров секвенирования функцией trim_paired (adapter_f, adapter_r) (последовательности адаптеров AmpliSeq были взяты из протокола 16S Metagenomic Sequencing Library Preparation); 3) объединение парных чтений при помощи merge_pairs (minlen = 150, maxns = 5, maxee = 2); 4) фильтрация по качеству с использованием функции q_score (min_quality = 20, max_ambiguous = 2); 5) дерепликация функцией dereplicate_sequences; 6) кластеризация до OTU с целью увеличения производительности программы (Operational Taxonomic Units) функцией claster_features_de_novo (perc_identity = 0.99); 7) удаление химер в uchime_ref; 8) идентификация организмов, где каждая OTU классифицируется на таксономический уровень, была произведена при помощи classify_sklearn(classifier), где classifier – это предварительно обученный на таксономической базе UNITE (version 9.0 2022.10.16) [2] (наивный байесовский классификатор fit_classifier_naive_bayes (reference_reads, reference_taxonomy)); 9) получение набора данных в виде двух архивов формата .qza (таблица частот и последовательности ДНК таксонов), пригодном для дальнейшего анализа разнообразия в QIIME 2.

Данный пайплайн был апробирован для анализа результатов секвенирования тотальной ДНК 210 образцов из проб торфа, растительного опада, древесины и микоризных корневых окончаний, отобранных в 2022 и 2023 годах на верховом болоте Мухрино в окрестностях города Ханты-Мансийск с целью оценки

разнообразия грибов верховых болот. ДНК выделяли набором SileksMagNA. Для метабаркодинга использовали регион ITS2 как наиболее информативный и традиционный для метабаркодинга грибов [3] и обеспеченный курируемыми референсными базами, такими как, например, UNITE. Секвенировали участок ITS2 в границах праймеров ITS7 и ITS4 [4, 5] по протоколу 16S Metagenomic Sequencing Library Preparation [6] на платформе Illumina MiSeq. Были получены сырые данные с разбросом от 1,5К до 250К прочтений на 1 пробу ДНК, из которых были отброшены пробы с числом прочтений < 10000. Сырые чтения (архивы FastQ и таблица метаданных) были загружены в NCBI Sequence Reads (номер биопроекта PRJNA1007262) и представлены в виде обработанной матрицы OTU на портале GBIF [7].

Результаты: Анализ последовательностей выявил 1609 операционных таксономических единиц (OTUs), классифицированных до 686 видов, 496 родов, 246 семейств, 89 отрядов, 33 классов, 9 типов и одного царства с уровнем сходства 99 %. Около 47 % таксонов идентифицированы на уровне вида, 22 % – на уровне рода, остальные – на более высоких таксономических уровнях. Основные порядки: Leotiomycetes, Dothideomycetes, Eurotiomycetes и др. (Ascomycota) и Agaricomycetes, Tremellomycetes, Mycrobotriomycetes и Exobasidiomycetes, (Basidiomycota).

Анализ структуры сообществ, выполненный для всех проб методом ординации, выявил четыре выделяющихся облака точек, соответствующих основным типам субстратов: поверхностный торф, древесина, опад растений, и глубинный торф. В свою очередь, ординация проб из каждой субстратной группы показала дробление в подгруппы по другим параметрам среды (например, пробы из опада растений – по видам хозяев). Полученные результаты свидетельствуют о достоверности полученных данных о видовом разнообразии и структуре сообществ грибов и корректной работе пайплайна.

Для проверки качества работы пайплайна также была произведена ручная классификация последовательностей грибов-макромицетов. Данную группу грибов предварительно изучали на участке Мухрино в течение 9 лет на постоянных площадках; была создана база последовательностей ITS и опубликован список видов [8]. При помощи BLAST [9] метабаркодинговые последовательности сравнивали с базой ITS плодовых тел видов, собранных на данном участке, и последовательностями ITS типовых образцов, представленных в NCBI. Последовательности были сгруппированы и проанализированы по семействам Agaricaceae, Auriscalpiaceae, Boletaceae, Clavariaceae, Cortinariaceae, Crepidotaceae, Entolomataceae, Hygrophoraceae, Inocybaceae, Lycoperdaceae, Lyophyllaceae, Mycenaceae, Omphalotaceae, Paxillaceae, Physalacriaceae, Pluteaceae, Psathyrellaceae, Russulaceae, Strophariaceae, Suillaceae, Thelephoraceae, Tricholomataceae, принадлежащих к трем основным порядкам Agaricales, Boletales и Russulales.

Результаты ручной классификации полностью совпали с полученными посредством пайплайна на уровне класса, порядка, семейства и рода. Однако на видовом уровне 23 % (27 из 118 видов) получили другие названия. В результате изучения проб грунта и опада список макромицетов на модельном участке верхового болота после метабаркодинга составил 161 вид. Выявленное разнообразие увеличилось на 66 видов (40 % от общего числа видов).

Выводы: подготовленный пайплайн позволяет корректно проанализировать секвенированные данные и с уверенностью определить до рода большинство последовательностей ITS2 грибов, полученных из проб грунта и растительных остатков. Классификация до видового уровня может успешно осуществляться с использованием ручной проверки на небольших датасетах. Данный пайплайн является частью разрабатываемого программного обеспечения, которое позволит в перспективе анализировать метабаркодинговые данные до уровня структуры сообществ.

Финансирование: Субсидия из федерального бюджета на выполнение государственного задания «Молекулярно-генетические методы в изучении и оценке состояния биоразнообразия Северных регионов (FENG-2024-0003)» (№ 1023041300017-6-1.6.4 от 13.03.2024).

Pipeline for processing metabarcoding data of fungi from natural communities

Ishmanov T.[1]*, Zvyagina E.[1, 2], Rudykina E.[1], Filippov I.[1], Bulyonkova T.[1], Dobrynina A.[1], Filippova N.[1]

[1] *Yugra State University, Khanty-Mansiysk, Russia*

[2] *Moscow State University, Moscow, Russia*

* *ishmanov2003tagir@mail.ru*

Key words: metabarcoding; eDNA; mycobiota; DNA-derived-data; Mukhrino; Western Siberia

Motivation and Aim: The assessment of microbiome diversity in natural communities has undergone a significant breakthrough in the last decade with the development of metabarcoding methods. However, at the present stage, an important challenge of eDNA sequencing data processing is the need to optimize the set and operating parameters of individual processing modules for the characteristics of the NGS-data obtained (presence of adapter residues, primers, etc.) and the systematic group of barcoded organisms (plants, fungi, invertebrates, etc.) To use the Illumina NGS metabarcoding in analyzing the diversity structure of fungal communities in natural ecosystems, a pipeline was assembled and optimized.

Methods and Algorithms: The pipeline was prepared in Python 3 in the Artifact API for QIIME 2 [1], which allows modifying the parameters depending on the goals of the study and the quality of the raw data. Pipeline stages: 1) submission of fastq files; 2) search and removal of sequencing adapters using the trim_paired function (adapter_f, adapter_r) (AmpliSeq adapter sequences were taken from the 16S Metagenomic Sequencing Library Preparation protocol); 3) combining paired reads using merge_pairs (minlen = 150, maxns = 5, maxee = 2); 4) filtering by quality using the q_score function (min_quality = 20, max_ambiguous = 2); 5) dereplication with the dereplicate_sequences function; 6) clustering to OTUs (Operational Taxonomic Units) in order to increase the performance of the program using the function cluster_features_de_novo (perc_identity = 0.99); 7) removal of chimeras in uchime_ref; 8) identification of organisms, where each OTU is classified to a certain taxonomic level, was carried out using classify_sklearn(classifier), where the classifier was pre-trained on the UNITE taxonomic database (version 9.0 2022.10.16) [2] (naive Bayes classifier

fit_classifier_naive_bayes (reference_reads, reference_taxonomy)); 9) obtaining a data set in the form of two archives in .qza format (table of frequencies and DNA sequences of taxa), suitable for further analysis of diversity in QIIME 2.

This pipeline was used to analyze the results of total DNA sequencing of 210 samples from peat, plant litter, wood and mycorrhizal root endings. The samples were collected in 2022 and 2023 in the Mukhrino raised bog near the city of Khanty-Mansiysk. DNA was isolated using the SileksMagNA kit. We used the ITS2 region for metabarcoding as it is traditionally considered to be the most informative for metabarcoding of fungi [3], and provided with curated reference databases, such as UNITE. The ITS2 region was sequenced within the boundaries of the ITS7 and ITS4 primers [4, 5] according to the 16S Metagenomic Sequencing Library Preparation protocol [6] on the Illumina MiSeq platform. Raw data was obtained ranging from 1.5K to 250K reads per 1 DNA sample, from which samples with the number of reads < 10,000 were discarded. Raw reads (FastQ archives and metadata table) were uploaded to NCBI Sequence Reads (bioproject number PRJNA1007262) and presented as a processed OTU matrix in the GBIF portal [7].

Results: Sequence analysis identified 1609 operational taxonomic units (OTUs), classified into 686 species, 496 genera, 246 families, 89 orders, 33 classes, 9 phyla and one kingdom with a similarity level of 99 %. About 47 % of taxa have been identified to the species level, 22 % to the genus level, and the rest to higher taxonomic levels. The main orders revealed were the Leotiomycetes, Dothideomycetes, Eurotiomycetes, etc. (Ascomycota) and Agaricomycetes, Tremellomycetes, Mycrobotriomycetes and Exobasidiomycetes, (Basidiomycota). Analysis of the community structure, performed for all samples using the ordination method, revealed four prominent clouds of points corresponding to the main types of substrates: surface peat, wood, plant litter, and deep peat. In turn, samples from each group are combined into their own subgroups according to other environmental parameters (for example, samples from plant litter – by the host species). Our results indicate the reliability of the data obtained on the species diversity and structure of fungal communities and the correct operation of the pipeline. To check the quality of the pipeline, manual classification of sequences belonging to fleshy fungi was also performed. This group of fungi was previously studied at the Mukhrino site for 9 years on permanent sites, a reference database of ITS sequences from fungal basidiomata was created, and a list of species has been published [8]. We have compare the metabarcoding sequences with the reference database and Type specimens sequences using BLAST [9]. The main families Agaricaceae, Auriscalpiaceae, Boletaceae, Clavariaceae, Cortinariaceae, Crepidotaceae, Entolomataceae, Hygrophoraceae, Inocybaceae, Lycoperdaceae, Lyophyllaceae, Mycenaceae, Omphalotaceae, Paxillaceae, ae, Pluteaceae, Psathyrellaceae, Russulaceae, Strophariaceae, Suillaceae, Thelephoraceae, Tricholomataceae, belonging to the three main orders Agaricales, Boletales and Russulales were analyzed. The results of manual classification completely matched those obtained using the pipeline at the level of class, order, family and genus. However, at the species level, 23 % (27 of 118 species) received different names. As a result of studying soil and litter samples, the list of fleshy fungi in the model area of the raised bog was expanded by 66 species (40 % of the total list).

Conclusion: the prepared pipeline allows us to correctly analyze the sequenced data and confidently determine to the genus the majority of ITS2 sequences of fungi obtained from soil and plant residue samples. Species-level classification can be successfully accomplished using manual inspection on small datasets. This pipeline is a part of

currently developed software that will allow analyzing metabarcoding data down to the community structure level.

Funding: Subsidy from the federal budget for the implementation of the state task "Molecular genetic methods in the study and assessment of the state of biodiversity of the Northern regions (FENG-2024-0003)" (No. 1023041300017-6-1.6.4 dated 03/13/2024).

Список литературы/References

1. Bolyen E., Rideout J.R., Dillon M.R. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37:852-857
2. Abarenkov K., Nilsson R.H., Larsson K.-H. et al. The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. *Nucleic Acids Res*. 2023;52(D1):D791-D797. doi 10.1093/nar/gkad1039
3. Tedersoo L., Bahram M., Zinger L., Nilsson R.H., Kennedy P.G., Yang T., Anslan S., Mikryukov V. Best practices in metabarcoding of fungi: From experimental design to results. *Mol Ecol*. 2022;31(10):2769-2795. doi 10.1111/mec.16460
4. White T., Bruns T., Lee S., Taylor J., Innis M., Gelfand D., Sninsky J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR protocols: A guide to methods and applications. 1990;315-322. doi 10.1016/B978-0-12-372180-8.50042-1
5. Ihrmark K., Bödeker I.T., Cruz-Martinez K. et al. New primers to amplify the fungal ITS2 region--evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol Ecol*. 2012;82(3):666-677. doi 10.1111/j.1574-6941.2012.01437.x
6. Sinclair L., Osman O.A., Bertilsson S., Eiler A. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the illumina platform. *PLoS One*. 2015;10(2):e0116955. doi 10.1371/journal.pone.0116955
7. Filippova N., Zvyagina E., Rudykina E.A., Ishmanov T.F., Filippov I.V., Bulyonkova T.M., Dobrynina A.S. DNA-based occurrence dataset on peatland fungal communities studied by metabarcoding in north-western Siberia. *Biodivers Data J*. 2024;12:e119851. doi 10.3897/BDJ.12.e119851
8. Filippova N., Zvyagina E., Rudykina E., Dobrynina A., Bolshakov S. The diversity of macromycetes in peatlands: nine years of plot-based monitoring and barcoding in the raised bog "Mukhrino", West Siberia. *Biodivers Data J*. 2023;11:e105111. doi 10.3897/BDJ.11.e105111
9. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410. doi 10.1016/S0022-2836(05)80360-2

doi 10.18699/bgrs2024-8.2-04

Идентификация и характеристика новых некультивируемых линий бактерий и архей из глубинной подземной биосферы

Кадников В.[1]*, Белецкий А.[1], Марданов А.[1], Карначук О.[2], Равин Н.[1]

[1] *Институт биоинженерии, ФИЦ Биотехнологии РАН, Москва, Россия*

[2] *Томский государственный университет, Томск, Россия*

* *vkadnikov@bk.ru*

Ключевые слова: некультивируемые микроорганизмы; геном; экология микроорганизмов; микробное сообщество; метагеномика

Мотивация и цель: Микроорганизмы представляют собой наиболее многочисленную форму жизни, однако наши знания об их разнообразии и осуществляемых ими процессах весьма ограничены, поскольку обычно не более 1 % микроорганизмов из природных сообществ удается культивировать и охарактеризовать традиционными методами. О существовании остальных 99 % «некультивируемых» стало известно лишь с появлением молекулярных методов анализа. Так, если в «домолекулярный» период было описано всего 12 филумов прокариот, то в настоящее время известно уже 175 филумов бактерий и 19 филумов архей (по последней версии Genome taxonomy database), более половины, из которых не имеет культивируемых представителей.

Хотя микроорганизмы «некультивируемых» групп широко распространены в различных экосистемах и могут играть ключевые роли в важных биогеохимических процессах, о них практически ничего (метаболизм, роль в биосфере и т. д.) не известно. Основным инструментом изучения некультивируемых микроорганизмов является метагеномный анализ, позволяющий не только охарактеризовать их на геномном уровне, но и описать генетическое разнообразие, метаболизм микробного сообщества и осуществляемые им биогеохимические процессы.

Одним из наименее изученных микробных местообитаний на Земле являются глубинные подземные экосистемы, которые могут оставаться изолированными от поверхности на протяжении миллионов лет. Большинство микроорганизмов глубиной подземной биосферы принадлежит к некультивируемым линиям, многие из которых специфичны для подземных местообитаний. Целью данной работы была идентификация и характеристика микроорганизмов подземной биосферы с помощью метагеномного анализа.

Методы и алгоритмы: Объектами исследования были глубинные подземные воды Томской области (2–3 км), длительное время (миллионы лет) изолированные от поверхности и грязевые вулканы Керченско–Таманского региона. Была определена температура воды, pH, Eh и ее химический состав, а также состав растворенного в ней газа. Также определили изотопный состав метана с целью уточнения его биогенного или абиогенного происхождения.

Характеристику состава сообществ проводили с помощью высокопроизводительного секвенирования ПЦР фрагментов генов 16S рРНК,

амплифицированных из метагеномной ДНК с «универсальными» праймерами 341F и 806R. Секвенирование проводили на Illumina MiSeq (2х 300 нт). Анализ состава сообществ по последовательностям генов 16S рРНК проводили с использованием пакета программ QIIME.

Для метагеномного анализа использовали технологии секвенирования Illumina и Oxford Nanopore. Для получения основного массива данных использовали Illumina (HiSeq2500). Сборку полученных последовательностей в контиги проводили с помощью metaSPAdes и/или MEGAHIT. Чтения MinION вначале собирали в контиги *de novo* с использованием Flye v. 2.7. Затем последовательности контигов корректировали с помощью Pilon v.1.2.2 в результате картирования на них чтений Illumina с использованием Bowtie 2. Вторым вариантом была совместная сборка чтений Illumina и MinION в контиги с помощью metaSPAdes hybrid assembler. Наконец, собранные только из чтений Illumina контиги определенного MAG (metagenome-assembled genome) объединялись в более протяженные или даже в полный кольцевой геном с использованием длинных чтений MinION. Для этого чтения MinION картировали на входящие в MAG контиги с помощью BWA v.0.7.15. Затем с помощью программы Npscarf v.1.0 формировали скаффолды, пробелы между контигами заполняли с использованием последовательностей Illumina из графа сборки metaSPAdes. Выбор конкретного варианта сборки метагенома определялся его сложностью, составом, объемом полученных последовательностей и т. д. Для биннинга метагеномных контигов в MAG использовали программы MetaBAT и/или CONCOCT. Собранные MAG таксономически классифицировали с использованием Genome Taxonomy Database Toolkit (GTDB-Tk). Доли отдельных микроорганизмов в сообществе определяли по доле прочтений метагеномной ДНК, вошедших в соответствующий MAG. Поиск генов и аннотацию MAG проводили с использованием NCBI Prokaryotic Genome Annotation Pipeline и/или RAST server. Для филогенетического анализа с помощью GTDB-Tk v.0.3.2 в MAG были найдены однокопийные маркерные гены и построено множественного выравнивания их конкатенированных последовательностей из данного MAG и всех видов из Genome Taxonomy Database. Выбранную часть множественного выравнивания, созданного в GTDB-Tk, использовали для построения филогенетического дерева с помощью PhyML v.3.3.

Результаты: Мы изучили микробные сообщества глубинных термальных водоносных горизонтов в мезозойских осадочных породах Западно-Сибирского региона с помощью метагеномного подхода. Были исследованы географически удаленные водоносные горизонты, доступ к которым осуществлялся через нефтепоисковые скважины глубиной 2–3 км. Выявлено три типа микробных сообществ. Сообщества первого типа (скважины 3P и Vi-1) состояли в основном из хемолитоавтотрофов, – метаногенных архей (Methanothermobacter) и сульфат-редуцирующих Bacillota и/или Desulfobacterota. Второй тип сообществ (скважина 1-R) содержал сульфат-редуцирующие Bacillota и Desulfobacterota, а также различные некультивируемые линии филумов Chlorofexota, Ignavibacteria, Aminicenantia и Riflebacteria, археи отсутствовали. Анализ геномов Chlorofexota, Ignavibacteria, Riflebacteria и Aminicenantia показал, что они могут сбраживать углеводы, а некоторые способны к аэробному и/или анаэробному дыханию. Вероятно, захороненное органическое вещество мезозойских отложений является субстратом для их роста, а образуемые водород и ацетат используются сульфат-

редукторами. Микробное сообщество третьего типа (скважина 5P) включало примерно равные доли бактерий и метаногенных архей. Среди бактерий найдены органотрофные представители Bacillota, Chloroflexota, Ignavibacteria, Armatimonadota, Pseudomonadota, Bacteroidetota и Spirochaetota. Сульфат-редукторы составляли небольшую долю и были представлены ‘Ca. Kapabacteria’ и Thermodesulfovibrio. Анализ метагеномов позволил получить несколько сот геномов микроорганизмов, в том числе представителей «некультивируемых» классов/филумов Aminicenantia, Bipolaricaulota, Patescibacteria, Atribacteriota, Armatimonadota, Riflebacteria, WOR-3 и BRC1.

Исследовано три грязевых вулкана Керченский полуостров – сопка Андрусова (KV1), сопка Тищенко (KV2) и вулкан «принц Ольденбургский» (KV3). Вулканы представляют собой небольшие (до нескольких метров) конусообразные структуры, из которых на поверхность постоянно поступают грязевые массы и газы (в основном метан). Микробные сообщества образцов флюида, отобранных из грязевых вулканов, в целом были сходны между собой, хотя имелись важные отличия, как между вулканами, так и между образцами, отобранными на разной глубине. Археи составляли от 9 до 55 % (доля возрастает с глубиной) и в основном были представлены организмами цикла метана, хотя в верхних горизонтах вулканов KV1 и KV2 также в значительных количествах встречались Bathyarchaeia. Большинство архей относилась к линиям, осуществляющим анаэробное окисление метана (ANME). Археи группы ANME-1 были обнаружены в двух верхних горизонтах (20 и 70 см) образцов KV1 и KV2, причем в KV1 были найдены ANME-1a (1–3 %), а в KV2 – ANME-1b (около 2 %). Группа ANME-2a-2b была найдена только в KV2, где ее доля возрастала с глубиной с 2 до 14%. Наиболее многочисленной группой ANME в образцах KV1 и KV2 была Methanoperedenaceae (осуществляют окисление метана, сопряженное с восстановлением нитрата или марганца), причем их доля возрастала с глубиной до 51 % в KV1 и до 14 % в KV2. В вулкане KV3 среди метан-окисляющих архей на всех глубинах доминировала группа ANME-3, на которую приходилось 32–47 % всего сообщества. ANME-3 также были найдены в глубинных слоях в KV1 (1.2 %) и KV2 (7.4 %). Доля метаногенных архей была намного ниже, чем ANME, в значительных количествах метаногены присутствовали только в KV3, в котором Methanocalculus составлял 3–6 % сообщества. Среди бактерий наиболее многочисленными были филумы Desulfobacterota, Pseudomonadota, Bacillota, Bacteroidota и Campylobacterota. Desulfobacterota в основном относились к сульфат-редуцирующим линиям (Desulfuromonadales и др.), причем их доля была максимальной в верхних горизонтах KV1 (28 %). Сульфат-редукторы филумов Bacillota (Desulfotomaculia) и Nitrospirota (Thermodesulfovibrionia) также присутствовали, но в меньших количествах. Среди протеобактерий наоборот, преобладали серо-окисляющие линии, в частности, рода Thiomicrospira (до 8 %), Thioprofundum, Thioalkalivibrio и др. Также среди потенциальных сероокислителей были найдены Campylobacterota (2–9 % сообществ), родов Sulfurimonas, Sulfurospirillum, Sulfuricurvum и Sulfurovum. В целом, выявленный нами состав микробных сообществ грязевых вулканов показал преобладание хемолитоавтотрфных микроорганизмов – участников циклов серы и метана, причем среди последних преобладали ANME археи.

Финансирование: Данная работа была выполнена при поддержке Российского научного фонда № 22-14-00178.

Identification and characterization of new uncultured lineages of bacteria and archaea from the deep subsurface biosphere

Kadnikov V.[1]*, Beletsky A.[1], Mardanov A.[1], Karnachuk O.[2], Ravin N.[1]

[1] *Institute of Bioengineering, Federal Research Center of Biotechnology RAS, Moscow, Russia*

[2] *Tomsk State University, Tomsk, Russia*

* *vkadnikov@bk.ru*

Key words: uncultured microorganisms; genome; microbial ecology; microbial community; metagenomics

Motivation and Aim: Microorganisms are the most abundant form of life, however, our knowledge of their diversity and the processes they carry out is very limited, since usually no more than 1 % of microorganisms from natural communities can be cultivated and characterized by traditional methods. The existence of the remaining 99 % "uncultivable" became known only with the advent of molecular methods of analysis. Thus, if in the "pre-molecular" period only 12 phyla of prokaryotes were described, then at present 175 phyla of bacteria and 19 phyla of archaea are known (according to the latest version of the Genome taxonomy database), more than half of which have no cultivated representatives.

Although microorganisms of "uncultivated" groups are widespread in various ecosystems and can play key roles in important biogeochemical processes, practically nothing is known about them (metabolism, role in the biosphere, etc.). The main tool for studying uncultivated microorganisms is metagenomic analysis, which allows not only to characterize them at the genomic level, but also to describe genetic diversity, the metabolism of the microbial community and the biogeochemical processes carried out by it.

One of the least studied microbial habitats on Earth are deep subsurface ecosystems, which can remain isolated from the surface for millions of years. Most microorganisms deep in the underground biosphere belong to uncultivated lineages, many of which are specific to underground habitats. The goal of this work was to identify and characterize microorganisms in the subsurface biosphere using metagenomic analysis.

Methods and Algorithms: The objects of study were deep underground waters of the Tomsk region (2–3 km), isolated from the surface for a long time (millions of years), and mud volcanoes of the Kerch-Taman region. The water temperature, pH, Eh and its chemical composition, as well as the composition of the gas dissolved were determined. The isotopic composition of methane was also determined in order to clarify its biogenic or abiogenic origin.

Characterization of the community composition was carried out using high-throughput sequencing of 16S rRNA gene fragments amplified from metagenomic DNA with "universal" primers 341F and 806R. Sequencing was performed on Illumina MiSeq (2x300 nt). Analysis of community composition based on 16S rRNA gene sequences was carried out using the QIIME software package.

Illumina and Oxford Nanopore sequencing technologies were used for metagenomic analysis. Illumina (HiSeq2500) was used to obtain the main dataset. The obtained sequences were assembled into contigs using metaSPAdes and/or MEGAHIT. MinION reads were first assembled into contigs *de novo* using Flye v. 2.7. The contig sequences were then corrected using Pilon v.1.2.2 as a result of mapping Illumina reads to them

using Bowtie 2. The second option was to jointly assemble Illumina and MinION reads into contigs using the metaSPAdes hybrid assembler. Finally, contigs of a specific MAG (metagenome-assembled genome) assembled only from Illumina reads were assembled into longer ones or even into a complete circular genome using MinION long reads. For this purpose, MinION reads were mapped to MAG contigs using BWA v.0.7.15. Scaffolds were then formed using the Npscarf v.1.0 program, and the gaps between contigs were filled using Illumina sequences from the metaSPAdes assembly graph. The choice of a specific metagenome assembly option was determined by its complexity, composition, volume of obtained sequences, etc. MetaBAT and/or CONCOCT programs were used for binning metagenomic contigs into MAGs. The assembled MAGs were taxonomically classified using the Genome Taxonomy Database Toolkit (GTDB-Tk). The relative abundancies of individual strains in the community were determined by the proportion of metagenomic DNA reads included in the corresponding MAG. Gene searching and MAG annotation were performed using the NCBI Prokaryotic Genome Annotation Pipeline and/or RAST server. For phylogenetic analysis using GTDB-Tk v.0.3.2, single-copy marker genes were found in MAGs and a multiple alignment of their concatenated sequences from these MAGs and all species from the Genome Taxonomy Database was constructed. A selected portion of the multiple alignment generated in GTDB-Tk was used to construct a phylogenetic tree using PhyML v.3.3.

Results: We studied microbial communities of deep thermal aquifers in Mesozoic sedimentary rocks of the West Siberian region using a metagenomic approach. Geographically distant aquifers, accessed through 2–3 km deep oil exploration wells, were explored. Three types of microbial communities were identified. Communities of the first type (wells 3P and Vi-1) consisted mainly of chemolithoautotrophs, methanogenic archaea (Methanothermobacter) and sulfate-reducing Bacillota and/or Desulfobacterota. The second type of community (well 1-R) contained sulfate-reducing Bacillota and Desulfobacterota, as well as various uncultivated lineages of the Chlorofexota, Ignavibacteria, Aminicenantia and Riflebacteria; archaea were absent. Analysis of the genomes of Chlorofexota, Ignavibacteria, Riflebacteria and Aminicenantia showed that they can ferment carbohydrates, and some are capable of aerobic and/or anaerobic respiration. It is likely that the buried organic matter of Mesozoic sediments is a substrate for their growth, and the hydrogen and acetate produced are used by sulfate reducers. The microbial community of the third type (well 5P) included approximately equal proportions of bacteria and methanogenic archaea. Among the bacteria, organotrophic members of Bacillota, Chloroflexota, Ignavibacteria, Armatimonadota, Pseudomonadota, Bacteroidetota and Spirochaetota were found. Sulfate reducers accounted for a small proportion and were represented by ‘Ca. Kapabacteria’ and Thermodesulfovibrio. Analysis of metagenomes made it possible to obtain several hundred genomes of microorganisms, including representatives of the “uncultured” classes/phyla Aminicenantia, Bipolaricaulota, Patescibacteria, Atribacteriota, Armatimonadota, Riflebacteria, WOR-3 and BRC1.

Three mud volcanoes of the Kerch Peninsula were studied – Andrusov Hill (KV1), Tishchenko Hill (KV2) and Prince of Oldenburg Volcano (KV3). Volcanoes are small (up to several meters) cone-shaped structures, from which mud masses and gases (mainly methane) constantly flow to the surface. The microbial communities of fluid samples collected from mud volcanoes were generally similar to each other, although there were important differences both between volcanoes and between samples collected at different depths. Archaea constituted from 9 to 55 % (the proportion increases with

depth) and were mainly represented by organisms of the methane cycle, although Bathyarchaeia were also found in significant quantities in the upper horizons of volcanoes KV1 and KV2. Most archaea belonged to lineages performing anaerobic methane oxidation (ANME). Archaea of the ANME-1 group were found in the upper two horizons (20 and 70 cm) of samples KV1 and KV2, with ANME-1a (1–3 %) found in KV1 and ANME-1b (about 2 %) in KV2. The ANME-2a-2b group was found only in KV2, where its proportion increased with depth from 2 to 14 %. The most numerous ANME group in samples KV1 and KV2 were Methanoperedenaceae (carry out methane oxidation coupled with the reduction of nitrate or manganese), and their share increased with depth to 51 % in KV1 and up to 14 % in KV2. In the KV3 volcano, the ANME-3 group dominated the methane-oxidizing archaea at all depths, accounting for 32–47 % of the total community. ANME-3 was also found in deep layers in KV1 (1.2 %) and KV2 (7.4 %). The proportion of methanogenic archaea was much lower than ANME, with methanogens present in significant quantities only in KV3, in which Methanocalculus comprised 3–6 % of the community. Among the bacteria, the most abundant phyla were Desulfobacterota, Pseudomonadota, Bacillota, Bacteroidota and Campylobacterota. Desulfobacterota mainly belonged to sulfate-reducing lineages (Desulfuromonadales and others), and their proportion was maximum in the upper horizons of KV1 (28 %). Sulfate reducers of the phyla Bacillota (Desulfotomaculia) and Nitrospirota (Thermodesulfovibrionia) were also present, but in smaller quantities. Among the Pseudomonadota, on the contrary, sulfur-oxidizing lineages predominated, in particular, the genus Thiomicrospira (up to 8 %), Thioprofundum, Thioalkalivibrio, etc. Also among potential sulfur-oxidizers were found Campylobacterota (2–9 % of communities), the genera Sulfurimonas, Sulfurospirillum, Sulfuricurvum and Sulfurovum. In general, the composition of the microbial communities of mud volcanoes that we identified showed the predominance of chemolithoautotrophic microorganisms – participants in the sulfur and methane cycles, and ANME archaea predominated among the latter.

Funding: The work was supported by the Russian Science Foundation (No. 22-14-00178).

doi 10.18699/bgrs2024-8.2-05

Биоплёнки пищевых производств – антропогенный экотоп обитания микроорганизмов порчи пищи и патогенов

Николаев Ю.А.[1]*, Юшина Ю.К.[2], Журина М.В.[1], Плакунов В.К.[1], Марданов А.В.[1], Эль-Регистан Г.И.[1]

[1] *ФИЦ Биотехнологии РАН, Москва, Россия*

[2] *ФНЦ Пищевых систем РАН, Москва, Россия*

* *NikolaevYA@mail.ru*

Ключевые слова: пищевые производства; биоплёнки; свойства; состав; патогены; порча пищи

Актуальность и цель: Биопленочная контаминация, в том числе патогенными бактериями, пищевых производств служит постоянным источником микробного загрязнения, угрожая не только качеству и безопасности пищевых продуктов, но и приводя к болезням, вызываемым пищевыми патогенами, являясь потенциальной базой их распространения.

Заболевания, передающиеся от животных, составляют 60 % всех инфекционных заболеваний людей, 75 % возникающих заболеваний людей (данные ООН). До 80 % бактериальных инфекций в США напрямую связаны с пищевыми патогенами [1]. Показано, что порча мясной продукции часто связана именно с образованием биопленок [2]. Биопленки на зарубежных пищевых производствах достаточно хорошо исследованы [3], для предприятий пищевой промышленности РФ таких данных практически не было. Целью работы было исследовать наличие, структуру и состав биоплёночной микробной контаминации некоторых мясокомбинатов РФ.

Методы исследования: В работе использованы методы микробиологии, трансмиссионной электронной микроскопии, молекулярной экологии, биоинформатики.

Результаты: С поверхностей производственных помещений мясо- и птицеперерабатывающих комбинатов отобраны несколько десятков проб. В большинстве проб со стен, трапов, потолков, конвейера, колёс транспорта обнаружены биоплёнки, или в стадии формирования или зрелые, состоящие из бактерий и матрикса различного состава.

Было выявлено: разнообразие морфотипов клеток, характерное для смешанных биоплёнок; морфологическое сходство клеток в БП разных образцов и их микроколониальный рост; возрастная гетерогенность клеток в пределах одной микроколонии с сосуществованием вегетативных и автолизированных клеток, покоящихся форм и, возможно, клеток-персистеров; гетерогенность химической природы полимерного матрикса по данным окрашивания рутением красным.

Метагеномный анализ показал высокую таксономическую сходность микроорганизмов, доминирующих в разных образцах. Из обнаруженных бактерий (археи не выявлены), относящихся к 11 филумам, доминантными были представители Actinobacteria, Bacteroidetes, Firmientes, Proteobacteria. Среди биопленочных контаминантов образцов были обнаружены бактерии, вызывающие порчу мясных, рыбных и молочных продуктов и способные к биопленкообразованию (p.p. Pseudomonas, Flavobacterium, Arcobacter, Vagococcus,

Chryseobacterium, Carnobacterium и др.), а также оппортунистические патогенны человека и животных (р.р. Arcobacter, Corinobacteria, Kosura, Listeria и др.). На основе геномных данных выявлена устойчивость бактерий в биоплёнках к антибиотикам: спектиномицину, сульфаметоксазолу, хлорамфениколу, ампициллину, трипетоприму, амикацину, амоксицилину и тобрамицину.
Выводы: Микробные биоплёнки достаточно распространены на мясоперерабатывающих предприятиях и представляют собой особый антропогенный экотоп, резервуар бактерий порчи пищи и патогенов. Повышенная стрессоустойчивость бактерий в биоплёнках и наличие детерминант антибиотикоустойчивости у бактерий, их населяющих, свидетельствуют о важности их исследования и борьбы с ними на предприятиях пищевой отрасли.

Biofilms of food production are an anthropogenic ecotope for the habitat of food spoilage microorganisms and pathogens

Nikolaev Yu.A.[1]*, Yushina Yu.K.[2], Zhurina M.V.[1], Plakunov V.K.[1], Mardanov A.V.[1], El-Registan G.I.[1]

[1] *Federal Research Center of Biotechnology RAS, Moscow, Russia*

[2] *Federal Research Center of Food Systems RAS, Moscow, Russia*

* *NikolaevYA@mail.ru*

Key words: food production; biofilms; properties; composition, pathogens, food spoilage

Importance and aim: Biofilm contamination, including pathogenic bacteria, in food production serves as a constant source of microbial contamination, threatening not only the quality and safety of food products, but also leading to diseases caused by foodborne pathogens, being a potential source of their spread.
Diseases transmitted from animals account for 60 % of all infectious diseases in humans, 75 % of emerging diseases in humans (UN data). Up to 80 % of bacterial infections in the United States are directly related to foodborne pathogens [1]. It has been shown that spoilage of meat products is often associated with the formation of biofilms [2]. Biofilms in foreign food production facilities have been studied quite well [3]; for food industry enterprises in the Russian Federation there was practically no such data. The purpose of the work was to investigate the presence, structure and composition of biofilm microbial contamination in some meat processing plants in the Russian Federation.
Research methods: The work used methods of microbiology, transmission electron microscopy, molecular ecology, and bioinformatics.
Results: Several dozen samples were taken from the surfaces of production premises of meat and poultry processing plants. In most samples from walls, ladders, ceilings, conveyors, and transport wheels, biofilms were found, either in the formation stage or mature, consisting of bacteria and a matrix of various compositions.
It was revealed: a variety of cell morphotypes characteristic of mixed biofilms; morphological similarity of cells in BP of different samples and their microcolonial growth; age-related heterogeneity of cells within one microcolony with the coexistence of vegetative and autolyzed cells, dormant forms and, possibly, persister cells; heterogeneity of the chemical nature of the polymer matrix according to ruthenium red staining.

Metagenomic analysis showed high taxonomic similarity of microorganisms dominant in different samples. Of the detected bacteria (archaea were not identified) belonging to 11 phyla, representatives of Actinobacteria, Bacteroidetes, Firmientes, and Proteobacteria were dominant. Among the biofilm contaminants of the samples, bacteria were found that cause spoilage of meat, fish and dairy products and are capable of biofilm formation (p.p. Pseudomonas, Flavobacterium, Arcobacter, Vagococcus, Chryseobacterium, Carnobacterium, etc.), as well as opportunistic pathogens of humans and animals (p.p. Arcobacter, Corinobacteria, Kosura, Listeria, etc.). Based on genomic data, the resistance of bacteria in biofilms to antibiotics was revealed: spectinomycin, sulfamethoxazole, chloramphenicol, ampicillin, tripetoprim, amikacin, amoxicillin and tobramycin.

Conclusion: Microbial biofilms are quite common in meat processing plants and represent a special anthropogenic ecotope, a reservoir of food spoilage bacteria and pathogens. The increased stress resistance of bacteria in biofilms and the presence of antibiotic resistance determinants in the bacteria inhabiting them indicate the importance of their study and control in the food industry.

Список литературы/References

1. Srey S., Jahid I.K., Ha S-D. Biofilm formation in food industries: A food safety concern. *Food Control.* 2013;31(2):572-585
2. Wang H., Qi J., Dong Y., Li Y., Xu X., Zhou G. Characterization of attachment and biofilm formation by meat-borne Enterobacteriaceae strains associated with spoilage. *LWT. Food Sci Technol*. 2017;86:399-407
3. Yang X., Wang H., He A., Tran F. Biofilm formation and susceptibility to biocides of recurring and transient Escherichia coli isolated from meat fabrication equipment. *Food Control*. 2018;90:205-211

doi 10.18699/bgrs2024-8.2-06

Действие различных физико-химических факторов на активность фосфат-аккумулирующих бактерий в лабораторных биореакторах, имитирующих промышленные очистные сооружения

Пименов Н.В.[1]*, Пелевина А.В.[1], Дорофеев А.Г.[1], Берестовская Ю.Ю.[1], Груздев Е.В.[2], Марданов А.В.[2]

[1] *ИНМИ, ФИЦ Биотехнологии РАН, Москва, Россия*

[2] *ИНБ, ФИЦ Биотехнологии РАН, Москва, Россия*

* *npimenov@mail.ru*

Ключевые слова: лабораторные биореакторы; активный ил очистных сооружений; фосфат-аккумулирующие бактерии; циклический метаболизм; гранулообразование; 16S рРНК профилирование микробного сообщества; анализ полных геномов

Процесс биогенного удаления фосфор известен с середины прошлого столетия и широко используется в биотехнологиях очистки сточных вод от соединений фосфора. Но до сих пор основные представители фосфатаккумулирующих бактерий (ФАО) не выделены в чистую культуру. Возможно, это связано с особенностями метаболизма ФАО, для развития которых необходимо чередовании анаэробной и аэробной фаз культивирования (циклический метаболизм). На очистных сооружениях динамика формирования ФАО-сообщества и его активность обусловлены конфигурацией и режимом работы реакторов, а также количеством и составом органического вещества. В природных экосистемах ФАО обычно обнаруживают на границе окисленных и восстановленных условий, где также происходит быстрая смена кислородного и бескислородного режимов. Выяснение особенностей формирования микробного сообщества ФАО, спектра используемых субстратов и путей их метаболизма позволит определить функциональные особенности ФАО, их место в микробном сообществе очистных сооружений, что может способствовать улучшению технологии удаления фосфора. Целью работы было выяснение особенностей гранулообразования, сукцессии ФАО-сообществ, спектра и путей метаболизма используемых ими органических субстратов.

Образцы активного ила для загрузки биореакторов были отобраны из аэробного реактора (аэротенка) блока удаления биогенных элементов Люберецких очистных сооружений (ЛОС, г. Москва) в котором реализована технология Кейптаунского университета (UCT), основанная на чередовании анаэробного и аэробного периодов культивирования.

Нами был создан лабораторный биореактор для циклического культивирования микробного сообщества на основе биореактора BIOSTAT B (фирмы “SARTORIUS”), имеющего рабочий объемом 2 л, снабженного перемешивающим устройством и внешней рубашкой для термостатирования. Цикличность культивирования заключалась в чередовании анаэробных условий с присутствием ацетата (легкодоступного источника углерода и энергии) и аэробных условий без ацетата (после его потребления в анаэробный период). Аэробные и анаэробные

условия создавали за счет подачи в биореактор воздуха или азота, очищенного от следов кислорода, с использованием системы регулирования подачи газов (ООО «ЭЛТОЧПРИБОР», РФ). Управление газовыми потоками и перистальтическими насосами осуществляли в автоматическом режиме с использованием универсального логического модуля LOGO (SIEMENS). В процессе длительного культивирования сообщества ФАО осуществляли непрерывный контроль физико-химических параметров, морфологических изменений структуры сообщества с использованием световой и электронной микроскопии с рентгеновским микроанализом, таксономического состава методом профилирования гена 16S рРНК, а также осуществили сборку полных геномов доминирующих микроорганизмов в составе сообщества ФАО.

В процессе культивирования ФАО сообщества в лабораторном биореакторе выявлено спонтанное образование гранулоподобных агрегатов разной морфологии. Электронно-микроскопические и молекулярные исследования закономерностей формирования и таксономического состава микроорганизмов гранулоподобных структур свидетельствует о том, что гранулообразование – естественный этап развития ФАО-сообщества. Таким образом отпадает необходимость в специфических технологиях получения гранулированного активного ила, что позволяет существенно упростить технологический процесс и снизить эксплуатационные затраты на процесс очистки сточных вод от фосфора.

В сравнении с ацетатом проведены исследования возможности использования ФАО-сообществом, обогащенным Ca. Accumulibacter, 17 органических соединений, которые наряду с ацетатом присутствуют в сточных водах. Выявлено их влияние на цикл выделения/поглощения фосфатов при смене анаэробных и аэробных периодов. В частности, установлено, что наряду с ацетатом, пропионатом и пируватом, полноценным субстратом для ФАО являются бутират, аспарагиновая и глутаминовая кислоты. С помощью метагеномного подхода были описаны основные метаболические пути Ca. Accumulibacter. Результаты работы могут быть использованы для разработки новых эффективных технологий очистки сточных вод от фосфора в России.

Финансирование: Работа выполнена при финансовой поддержке РНФ № 21-64-00019 и Министерства науки и высшего образования Российской Федерации.

Effect of physicochemical factors on activity of phosphate-accumulating bacteria in laboratory bioreactor models of industrial waste treatment facilities

Pimenov N.V.[1]*, Pelevina A.V.[1], Dorofeev A.G.[1], Berestovskaya Yu.Yu.[1], Grouzdev E.V.[2], Mardanov A.V.[2]

[1] *Winogradsky Institute of Microbiology, Federal Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russia*

[2] *Institute of Bioengineering, Federal Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russia*

** npimenov@mail.ru*

Key words: laboratory bioreactors; activated sludge of waste treatment facilities; phosphate-accumulating bacteria; cyclic metabolism; granule formation; 16S rRNA profiling of microbial communities; complete genome analysis

Biological phosphorus removal has been known since the mid-20th century and is widely used in biotechnologies for phosphorus removal from wastewater. The major representatives of phosphate-accumulating organisms (PAO) have not been, however, obtained as pure cultures. This is probably due to the properties of PAO metabolism, since development of these bacteria requires alternations of the oxic and anoxic phases of cultivation (cyclic metabolism). Dynamics of formation and activity of the PAO communities of waste treatment plants are determined by the configuration and operation mode of their bioreactors and by the amount and composition of organic matter. In natural environments, PAO are usually detected at the redox interface, where rapid shifts between oxic and anoxic conditions also occur. Determination of the characteristics of formation of the PAO microbial communities, the spectrum of utilized substrates, and the pathways of their metabolism will result in elucidation of the PAO functional properties and their place in the communities of waste treatment facilities, which may help in improvement of the phosphorus removal technology. The goal of the present work was to determine the patterns of granule formation, succession of the PAO communities, and the spectrum of utilized substrates and the relevant metabolic pathways.

The bioreactors were loaded with activated sludge samples collected from an aerobic digester (aerotank) of the block for biogenic elements removal at the Lyubertsy waste treatment plant (Moscow), in which the University of Cape Town (UCT) technology based on alternation of aerobic and anaerobic cultivation was implemented.

The laboratory bioreactor for cyclic cultivation of the microbial community was constructed by modification of a BIOSTAT B reactor (Sartorius) with 2-L working volume, equipped with an agitator and an external thermostat jacket. Cyclic cultivation was achieved by switching from anoxic conditions with acetate as an easily available source of carbon and energy to oxic conditions without acetate (after its consumption during the anoxic phase). Oxic and anoxic conditions were established by bubbling the bioreactor with air or oxygen-free nitrogen, respectively by means of the regulated gas supply system (Eltochpribor, Russia). Gas flows and the operation of peristaltic pumps were controlled automatically using the LOGO universal logical module (Siemens). In the course of long-term cultivation of PAO communities, the physicochemical parameters were monitored continuously. Morphological changes in the community structure were studied by light microscopy and electron microscopy coupled to X-ray microanalysis. The taxonomic composition of the community was determined by the 16S rRNA gene profiling. Whole genomes of the dominant members of the PAO microbial community were assembled.

During cultivation of the PAO community in a laboratory bioreactor, spontaneous formation of granule-like aggregates of diverse morphology was observed. Electron microscopic and molecular investigation of formation patterns of the granule-like structures and of their taxonomic composition showed granule formation to be a natural stage of development for PAO communities. Thus, since no specific technologies to obtain granulated activated sludge are required, the technological process may be simplified considerably, with decreased operating costs for phosphorus removal from wastewater.

Apart from acetate, the possible utilization of 17 organic compounds also occurring in wastewater by the PAO community enriched with *Ca*. Accumulibacter was investigated. The effect of these compounds on the phosphate release/consumption cycle during alternating oxic and anoxic phases was studied. Thus, it was found that, apart from acetate, propionate, and pyruvate, butyrate, aspartate, and glutamate were adequate substrates for PAO. The metagenomic approach was used to describe the main metabolic pathways of *Ca*. Accumulibacter.

The results of this work may be used to develop new efficient technologies for phosphorus removal from wastewater at Russian waste treatment plants.

Funding: The study was supported by the Russian Science Foundation (No. 21-64-00019) and by the Russian Federation Ministry of Science and Higher Education.

doi 10.18699/bgrs2024-8.2-07

Состав и метаболический потенциал микробных сообществ пещеры Шульган-Таш: сравнение метабаркодинга и метагеномики

Полякова Е.Ю.[1]*, Балкин А.С.[1, 2], Шагимарданова Е.И.[1], Кузьмина Л.А.[3], Червяцова О.Я.[4], Гоголев Ю.В.[1, 5], Гоголева Н.Е.[1, 2, 6]

[1] *Институт фундаментальной медицины и биологии, Казанский федеральный университет, Казань, Россия*

[2] *Институт клеточного и внутриклеточного симбиоза Уральского отделения Российской академии наук, Оренбург, Россия*

[3] *Уфимский институт биологии, Уфимский федеральный исследовательский центр, Российская академия наук, Уфа, Россия*

[4] *Государственный природный заповедник «Шульган-Таш», Иргизлы, Россия*

[5] *Казанский институт биохимии и биофизики, Федеральный исследовательский центр «Казанский научный центр Российской академии наук», Казань, Россия*

[6] *Научно-исследовательский отдел лимнологии, Университет Инсбрука, Мондзее, Австрия*

* *polyakova.e.yu@gmail.com*

Ключевые слова: метабаркодинг; метагеном; биопленки

Мотивация и цель: В современной микробной экологии сформировались два основных подхода для изучения микробиомов: метагеномика, основанная на полногеномном секвенировании и метабаркодирование. Принято считать, что второй подход, основанный на секвенировании отдельных таксономически значимых последовательностей уступает в релевантности моделирования метаболического потенциала сообщества в сравнении с метагеномным анализом [1]. В то же время метабаркодинг является более доступным с точки зрения финансовых и трудовых затрат, а появление новых инструментов значительно расширяет его возможности. В связи с этим, нами была поставлена задача по сравнению результатов исследований с использованием двух подходов на примере микробных сообществ пещеры Шульган-Таш. Объектом наших исследований является пещера Шульган-Таш, расположенная в предгорьях Южного Урала. Пещеры представляют собой олиготрофные экосистемы, где преобладают умеренно низкие температуры и высокая влажность воздуха, а отсутствие света накладывает ограничения на способ получения энергии. Такие экстремальные условия обитания предполагает ограниченное количество участников сообщества, что делает их удобной модельной системой.

Методы и алгоритмы: Для сравнительных исследований были выбраны настенная биопленка «Азурит» и погруженная водная биопленка из пещерного озера, которые значительно отличались друг от друга по составу и метаболическим характеристикам. Метод и результаты метабаркодинга настенной биопленки были опубликованы ранее [2]. Биоинформатический анализ данных шотган-секвенирования был выполнен по следующей схеме: удаление последовательностей адаптеров и фильтрация ридов по качеству с помощью Trimmomatic, сборка метагенома с использованием metaspades, бининг, доработка

и повторная сборка бинов в пакете metawrap. Качество Metagenome-assembled genomes (MAGs) на всех этапах проверяли с использованием checkM. Дополнительно MAGs фильтровались программой refinem. Обилие каждого MAG в метагеноме высчитывали программой CoverM. Таксономическое профилирование делали с использованием базы данных GTDB r214.

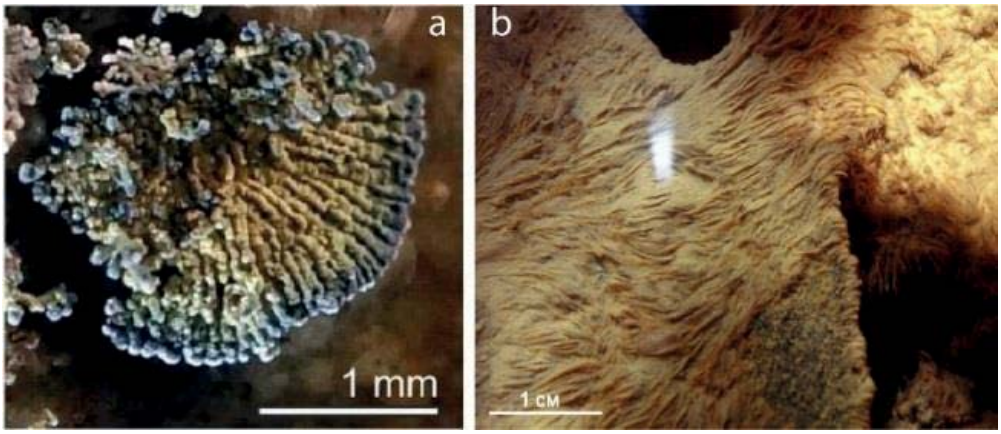


Рис. 1. Биопленки пещеры Шульган-Таш: настенное обрастание Азурит (*a*), погруженная водная биопленка пещерного озера «Дальнее» (*b*)

Результаты: Биопленка «Азурит» по данным метабаркодинга была представлена филами Actinomycetota (> 70 % сообщества) и Pseudomonadota (> 13 %), а также в меньшей степени филами Acidobacteriota (~3 %), Planctomycetota (~2.5 %), Chloroflexota (~1 %). На более низком таксономическом уровне доминировали неидентифицированный представитель семейства Pseudonocardiaceae (~62 %), некультивируемый представитель рода (MAG) Palsa-1315 (семейство Nitrospiraceae), g. Pseudonocardia, некультивируемый род (MAG) WHTC01 (семейство Egibacteraceae), неидентифицированный представитель класса Gammaproteobacteria, а также представители рода Miltoncostaea, ранее обнаруженный в морских местообитаниях.

В водной биопленке преобладали представители Pseudomonadota, Acidobacteriota, Planctomycetota, Methylomirabilota, Chloroflexota, и Nitrospirota, а также Actinomycetota, причем доля последних составляла немногим более 5 %. Наиболее представленные таксоны на уровне рода классифицировались как ранее собранные MAG, а именно Palsa-1315 (семейство Nitrospiraceae), DSQQ01 (класс Anaerolineae), AR19 (порядок Rokubacteriales), CSP1-1 (семейство Nitrosopumilaceae), 40CM-68-15 (порядок Rokubacteriales), GMQP-bins7 (семейство Gaiellaceae), QEVD01, Gp6-AA45 (порядок Vicinamibacterales), VGXF01 (семейство Nitrospiraceae).

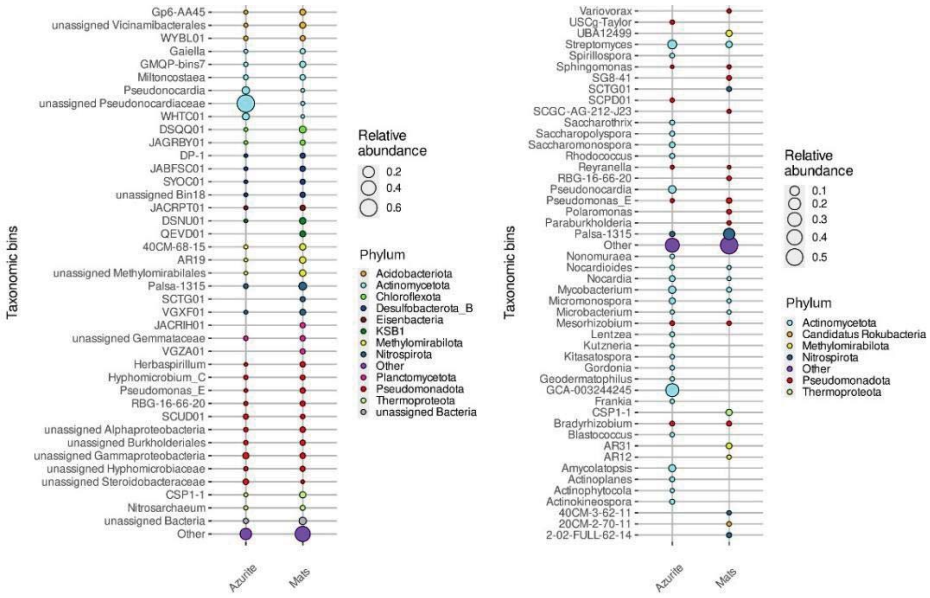


Рис. 2. Наиболее представленные рода бактерий в образцах «Азурит» и водной биопленке, выявленных с помощью 16S метабаркодинга (слева) и shotgun-секвенирования метагенома (справа). Размер круга отражает относительную численность ампликонов (ASV для данных метабаркодинга), либо покрытие ридов (для образцов shotgun) в образцах, цветом круга обозначены филы бактерий

Результаты метагеномного анализа биопленок были сходны с результатами метабаркодинга. В настенной биопленке преобладали микроорганизмы из филы *Actinomycetota* (более 70 % сообщества), далее филы *Pseudomonadota* (~16 %) и *Nitrospirota* (~1 %). Доминантом, как и в метабаркодинге, оказалась некультивируемая бактерия семейства Pseudonocardiaceae. Далее следовали представители родов *Streptomyces*, *Pseudonocardia, Mycobacterium, Amycolatopsis*. MAG Palsa-1315 (семейство *Nitrospiraceae*) составлял более 1 % сообщества. В водной биопленке по результатам шотган-секвенирования преобладали те же филы, что и по результатам метабаркодинга, но доли фил незначительно отличались, а актинобактерии составляли ~14 % сообщества. MAG Palsa-1315 являлся доминантом.

Выводы: Метод метабаркодинга для быстрого и недорого скрининга образцов являлся релевантным для биопленок олиготрофных местообитаний. Сравнение данных метабаркодинга и метагеномики демонстрировали сходные результаты на уровне доминант, но отличались по обилию минорных представителей биопленок. Значительная часть микробного сообщества не идентифицировалась до уровня рода обоими методами. Метагеномный анализ являлся предпочтительным для характеристики метаболизма микробного сообщества.

Финансирование: Исследование поддержано Программой стратегического академического лидерства Казанского федерального университета (ПРИОРИТЕТ-2030). Работа поддержана грантом РНФ 22-14-00317.

Microbial community diversity and metabolic potential in the Shulgan-Tash Cave: insights from metabarcoding and metagenomics

Polyakova E.[1]*, Balkin A.[1, 2], Shagimardanova E.[1], Kuzmina L.[3], Chervyatsova O.[4], Gogolev Y.[1, 5], Gogoleva N.[1, 2, 6]

[1] *Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[2] *Institute for Cellular and Intracellular Symbiosis, Ural Branch of the Russian Academy of Sciences, Orenburg, Russia*

[3] *Ufa Institute of Biology, Ufa Federal Research Center, Russian Academy of Sciences, Ufa, Russia*

[4] *State Nature Reserve "Shulgan-Tash", Irgyzly, Russia*

[5] *Kazan Institute of Biochemistry and Biophysics, Federal Research Center "Kazan Scientific Center of the Russian Academy of Sciences", Kazan, Russia*

[6] *Research Department for Limnology, Mondsee, Universität Innsbruck, Mondsee, Austria*

Key words: metabarcoding; metagenome; biofilm

Motivation and Aim: In modern microbial ecology, two main approaches for studying microbiomes have emerged: metagenomics, based on whole-genome sequencing, and metabarcoding. It is generally believed that metabarcoding, which involves sequencing individual taxonomically significant sequences, is inferior to metagenomic analysis in modeling a community's metabolic potential. However, metabarcoding is more accessible in terms of financial and labor costs, and the emergence of new tools significantly expands its capabilities. In this context, we aimed to compare the results of studies using these two approaches by examining the microbial communities of Shulgan-Tash Cave, located in the foothills of the Southern Urals. Caves are oligotrophic ecosystems characterized by moderately low temperatures, high humidity, and the absence of light, which imposes restrictions on energy production. These extreme habitat conditions result in a limited number of community members, making caves a convenient model system.

Methods and Algorithms: For our comparative studies, we selected two biofilms: an azurite wall biofilm and a submerged aquatic biofilm from a cave lake, which differed significantly in composition and metabolic characteristics. Previous research has published the metabarcoding results of the wall biofilm [1]. For metagenomic analysis, we used a comprehensive bioinformatics pipeline: removal of adapter sequences and quality filtering with Trimmomatic, metagenome assembly with metaSPAdes, binning, refining, and re-assembly of bins using the MetaWRAP package. The quality of metagenome-assembled genomes (MAGs) was checked using CheckM, and MAGs were further refined with RefineM. The abundance of each MAG in the metagenome was calculated with CoverM, and taxonomic profiling was conducted using the GTDB r214 database.

Results: Metabarcoding data indicated that the azurite biofilm was predominantly composed of Actinomycetota (> 70 % of the community) and Pseudomonadota (> 13 %), with smaller contributions from Acidobacteriota (~3 %), Planctomycetota (~2.5 %), and Chloroflexota (~1 %). The dominant taxa included an unidentified representative of the family Pseudonocardiaceae (~62 %), the uncultivated genus Palsa-1315 (family Nitrospiraceae), Pseudonocardia, WHTC01 (family Egibacteraceae), an unidentified

member of the class Gammaproteobacteria, and members of the genus Miltoncostaea, previously found in marine habitats.

The aquatic biofilm was dominated by Pseudomonadota, Acidobacteriota, Planctomycetota, Methylomirabilota, Chloroflexota, and Nitrospirota, with Actinomycetota accounting for just over 5 % of the total. The most represented genera included Palsa-1315 (family Nitrospiraceae), DSQQ01 (class Anaerolineae), AR19 (order Rokubacteriales), CSP1-1 (family Nitrosopumilaceae), 40CM-68-15 (order Rokubacteriales), GMQP-bins7 (family Gaiellaceae), QEVD01, and Gp6-AA45 (order Vicinamibacterales).

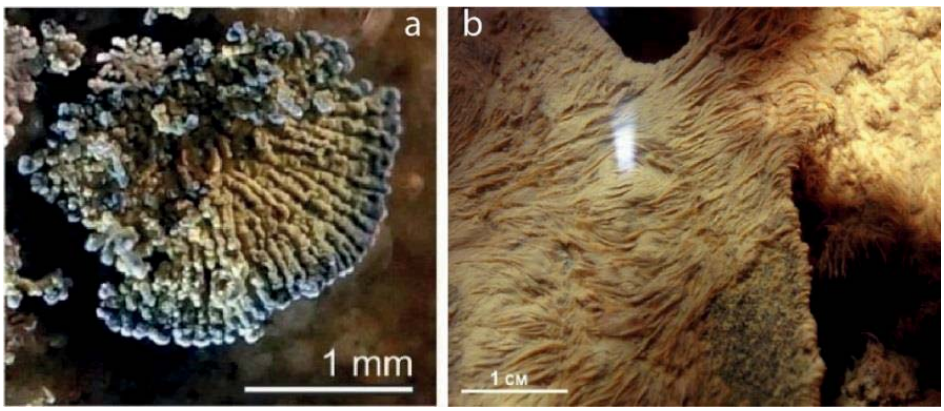


Fig. 1. Biofilms of Shulgan-Tash cave: azurite wall biofilm (*a*), submerged aquatic biofilm from a cave lake (*b*)

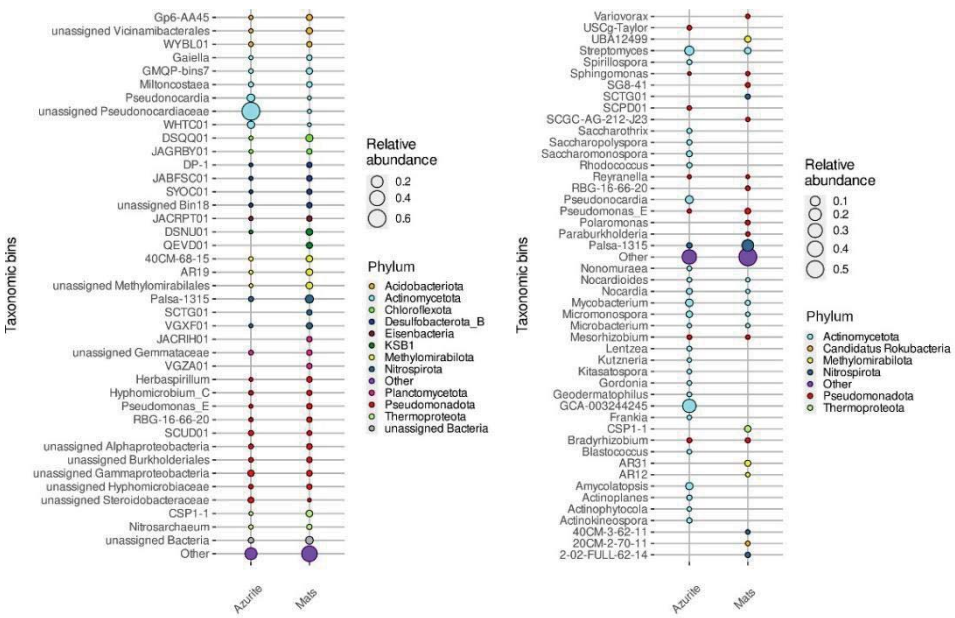


Fig. 2. The most represented bacterial genera in azurite and aquatic biofilm samples identified by 16S metabarcoding (left) and shotgun sequencing (right). The size of the circle reflects the relative abundance of amplicons (ASV for metabarcoding data) or reed coverage (for shotgun samples) in the samples; the color of the circle indicates bacterial phyla

The results of metagenomic analysis were consistent with those of metabarcoding. The wall biofilm was dominated by Actinomycetota (> 70 % of the community), followed by Pseudomonadota (~16 %) and Nitrospirota (~1 %). The dominant bacterium was an uncultivated member of the family Pseudonocardiaceae, followed by representatives of the genera Streptomyces, Pseudonocardia, Mycobacterium, and Amycolatopsis. The MAG Palsa-1315 (family Nitrospiraceae) accounted for more than 1 % of the community. The aquatic biofilm also showed similar dominant phyla as the metabarcoding results, with slight differences in proportions and Actinobacteria comprising ~14 % of the community. MAG Palsa-1315 was the dominant species.

Conclusion: The metabarcoding method for rapid and inexpensive screening of samples was relevant for biofilms in oligotrophic habitats. Comparison of metabarcoding and metagenomics data showed similar results at the level of dominant taxa but differed in the abundance of minor representatives of biofilms. The correspondence between metabarcoding and metagenome data depended on the database used. A significant portion of the microbial community was not identified to the genus level by either method. Metagenomic analysis has been preferred for characterizing microbial community metabolism.

Funding: The study is supported by Kazan Federal University (PRIORITY-2030) and by RSF project ID: 22-14-00317.

Список литературы/References

1. Rieder J., Kapopoulou A., Bank C. et al. Metagenomics and metabarcoding experimental choices and their impact on microbial community characterization in freshwater recirculating aquaculture systems. *Environ Microbiome.* 2023;18(1):8. doi 10.1186/s40793-023-00459-z
2. Gogoleva N., Chervyatsova O., Balkin A. et al. Microbial tapestry of the Shulgan-Tash cave (Southern Ural, Russia): influences of environmental factors on the taxonomic composition of the cave biofilms. *Environ Microbiome.* 2023;18(1):82. doi 10.1186/s40793-023-00538-1

doi 10.18699/bgrs2024-8.2-08

Ацетогенные прокариоты с новыми механизмами запасания энергии

Фролов Е.Н.*, Гололобова А.В., Ельченинов А.Г., Лебединский А.В.
ФИЦ Биотехнологии РАН, Москва, Россия
** evgenii_frolov_89@mail.ru*

Ключевые слова: ацетогенез; восстановительный глициновый путь; полисульфидредуктаза; путь Вуда–Льюнгдаля

Мотивация и цель: Определение величины минимального энергетического выхода реакции, достаточного для использования клеткой (биологический квант энергии), является одной из важнейших проблем современной биологии. Объектом исследования выступают микроорганизмы, живущие за счёт биохимических реакций, дающих минимальный выход энергии, к которым относятся ацетогенные прокариоты, формиатотрофные гипертермофильные археи и синтрофные микроорганизмы, осуществляющие окисление спиртов и жирных кислот. Решение данной проблемы приведёт не только к пониманию границ распространения жизни на Земле, но и открытию новых метаболических процессов, позволяющих жить в условиях, близких к термодинамическому пределу. Решение поставленной проблемы имеет и важный практический аспект ввиду высокого биотехнологического потенциала ацетогенных прокариот, который обусловлен их уникальной способностью синтезировать *de novo* органические соединения из H_2/CO_2, CO или же синтез-газа. Целью работы является изучение новых метаболических процессов, влияющих на величину биологического кванта энергии у ацетогенных бактерий, а также исследование распространения выявленных метаболических процессов среди прокариот.
Методы и алгоритмы: Выделение и характеристика новых микроорганизмов была выполнена на модифицированной среде Пфеннига с применением стандартных микробиологических процедур. Определение полной последовательности геномов было выполнено с применением технологий секвенирования Illumina и Oxford Nanopore с последующей гибридной сборкой с помощью SPAdes, Unicycler и Canu. Первичная аннотация генома была выполнена с помощью NCBI Prokaryotic Genome Annotation Pipeline в процессе подачи генома на портал NCBI. Дополнительные аннотации, направленные на улучшение предсказания функции белков, были выполнены с помощью IMG/MER и веб-сервера RAST, а также поиска ключевых метаболических генов вручную. Поиск потенциальных ацетогенов, в том числе обладающих новыми механизмами запасания энергии, был проведен в общедоступных базах данных прокариотных геномов. Кроме того, для выполнения данной работы была создана локальная база данных на основе репрезентативных геномов GTDB (47 894 геномов, представляющих все виды прокариот), был организован необходимый для поисков функционал, и было исследовано распределение ключевых детерминант известных и гипотетических вариантов ацетогенеза.
Результаты: Из горячих источников полуострова Камчатка были выделены и охарактеризованы первые облигатно хемолитоавтотрофные ацетогенные

бактерии, штаммы 3443-3Ac и 3819-GS1, которые являются облигатными анаэробами, умеренными термофилами, растут при слабокислых и нейтральных значениях pH и не требуют наличия витаминов в среде. Штамм 3443-3Ac рос исключительно в хемолитоавтотрофных условиях в присутствии H_2/CO_2, CO или формиата, а конечным продуктом метаболизма являлся ацетат. В случае внесения в среду элементной серы и полисульфида помимо ацетата начинал образовывался сульфид. На основании полученных результатов физиологических и филогеномных исследований предложен новый род и вид *Aceticella autotrophica*.
Анализ генома позволил выявить гены, кодирующие полного набор белков автотрофного ацетогенеза, в котором окислительный модуль был представлен бифуркационной гидрогеназой, восстановительный – путём Вуда-Льюнгдаля (ПВЛ), а энергетический – ECH гидрогеназой и H+-зависимой АТФ-синтазой. Предложенные схемы метаболизма *Ac. autotrophica* 3443-3Ac^T при росте на H_2/CO_2, HCOOH и CO предполагают работу одного и того же ферментативного аппарата, различия заключаются в участии монофункциональной CO-дегидрогеназы при росте на CO и противоположном направлении работы H_2-зависимой CO_2-редуктазы при росте на формиате. Для *Ac. autotrophica* 3443-3Ac^T был предложен новый механизм запасания энергии – «облегчённый» ацетогенез, при котором катаболическую функцию ПВЛ, связанную с реокислением восстановительных эквивалентов, частично берет на себя НАД(Ф)H-полисульфидзависимая редуктаза. Увеличение количества окисленного НАД(Ф)$^+$ позволяет бифуркационной гидрогеназе более эффективно вырабатывать низкопотенциальные ферредоксины, необходимые для генерации мембранного потенциала ECH комплексом.
Штамм 3819-GS1 рос исключительно в хемолитоавтотрофных условиях с H_2 в качестве донора электронов и CO_2 в качестве акцептора электронов и источника углерода, а единственным продуктом метаболизма является ацетат. Новый изолят не был способен к восстановлению внешних акцепторов электронов. На основании филогеномных исследований предложен новый род и вид *Acetitalea autotrophica*. Анализ генома выявил существенные отличия метаболизма нового изолята от метаболизма всех известных на данный момент ацетогенных микроорганизмов. Было предположено, что окислительный модуль представлен бифуркационной гидрогеназой, восстановительный – восстановительным глициновым путём, а энергетический – 11-субъединичным Nuo-подобным мембранным комплексом и H^+-зависимой АТФ-синтазой. Таким образом, нами был предложен новый механизм «глицинового» ацетогенеза.
Были выработаны критерии отнесения микроорганизмов к потенциальным ацетогенам («классическим» и гипотетическим) и критерии, позволяющие судить о наличии или отсутствии у них известных для ацетогенов механизмов запасания энергии (Rnf-комплекс и ECH-гидрогеназа). Было выявлено 280 потенциальных «классических» ацетогенов в 18 филумах. Большинство относились к филумам Bacillota_A (178) и Bacillota_B-G (от 22 до 4) и было рассеяно внутри этих филумов, численно не преобладая ни в одной из ее линий. Эти результаты согласуются с представлением о вертикальном наследовании способности к «классическому» ацетогенезу и с нашим с выводом о вертикальном наследовании этой способности, сделанным на основании филогенетических деревьев ключевых детерминант. Был обнаружен 61 геном потенциальных необычных ацетогенов (ПВЛ при отсутствии Rnf и ECH). Однако в тех случаях, когда это были геномы

культивируемых микроорганизмов, они чаще всего принадлежали сульфатредукторам из порядков *Desulfotomaculales* и *Desulfitobacteriales*, и ПВЛ у них скорее всего выполняет иную чем ацетогенез функцию. Было обнаружено 187 геномов потенциальных «глициновых» ацетогенов, и из них всего в шести отсутствовали гены ECH и Rnf. Таким образом, паттерн детерминант, наблюдаемый у «глицинового» ацетогена *At. autotrophica* 3819-GS1, является весьма редким. Интересно, что из этих 187 потенциальных ацетогенов 27 имели детерминанты как глицинового пути, так и ПВЛ.
Выводы: Таким образом, в ходе проделанной работы были выделены и охарактеризованы первые облигатно хемолитоавтотрофные ацетогенные бактерии, *Aceticella autotrophica* и *Acetitalea autotrophica*. Для *Ac. autotrophica* 3443-3Ac^T предложен новый механизм запасания энергии – «облегчённый» ацетогенез, при котором катаболическую функцию ПВЛ, связанную с реокислением восстановительных эквивалентов, частично берёт на себя НАД(Ф)Н-зависимая полисульфидредуктаза. Для *At. autotrophica* предложен новый механизм «глицинового» ацетогенеза. Анализ геномных баз данных наличие новых механизмов запасания энергии также среди бактерий.
Финансирование: Исследование поддержано грантом РНФ № 24-14-00177.

Acetogenic prokaryotes with new energy conservation mechanisms

Frolov E.N.*, Gololobova A.V., Elcheninov A.G., Lebedinsky A.V.
Research Center of Biotechnology RAS, Moscow, Russian
** evgenii_frolov_89@mail.ru*

Key words: acetogenesis; reductive glycine pathway; polysulfide reductase; Wood–Ljungdahl pathway

Motivation and Aim: Determining the minimum energy yield of a reaction sufficient for use by a cell (biological quantum of energy) is one of the most important problems of modern biology. The objects of the study are microorganisms living due to biochemical reactions that provide minimal energy output. Such microorganisms include acetogenic prokaryotes, formatotrophic hyperthermophilic archaea, and syntrophic microorganisms that carry out the oxidation of alcohols and fatty acids. Solving this problem will lead not only to understanding the boundaries of the distribution of life on Earth, but also to the discovery of new metabolic processes that allow living in conditions close to the thermodynamic limit. Solving this problem also has an important practical aspect due to the high biotechnological potential of acetogenic prokaryotes, which is due to their unique ability to synthesize *de novo* organic compounds from H_2/CO_2, CO or synthesis gas. The aim of this work is to study new metabolic processes affecting the value of biological quantum of energy in acetogenic bacteria and to investigate the distribution of the identified metabolic processes among prokaryotes.
Methods and Algorithms: The isolation and characterization of new microorganisms was performed on modified Pfennig's medium using standard microbiological procedures. Genome sequencing was performed using Illumina and Oxford Nanopore sequencing technologies followed by hybrid assembly using SPAdes, Unicycler and Canu. Primary genome annotation was performed by Prokaryotic Genome Annotation Pipeline during the process of genome submission to NCBI submission portal. Additional annotations

aimed to improve predictions of protein function were performed using IMG/MER System and Rapid Annotation using Subsystem Technology web server and manual curation of key metabolic genes. In databases of prokaryotic genomes, searches were conducted for potential acetogens, including those that have new energy conservation mechanisms. Moreover, a local database was created based on representative GTDB genomes (47,894 genomes representing all species of prokaryotes), the functionalities necessary for the search were organized, and the distribution of key determinants of known and hypothetical variants of acetogenesis was investigated.

Results: The first obligate chemolithoautotrophic acetogenic bacteria, strains 3443-3Ac and 3819-GS1, were isolated from hot springs of Kamchatka peninsula and characterized. These are obligate anaerobes and moderate thermophiles that grow at slightly acidic and neutral pH values and do not require the presence of vitamins in the medium. Strain 3443-3Ac grows exclusively under chemolithoautotrophic conditions in the presence of H_2/CO_2, CO or formate. The end product of metabolism is acetate. It was also shown to be able to reduce elemental sulfur, polysulfide and thiosulfate, forming sulfide in parallel with acetate. Based on the results of physiological and phylogenomic studies, a new genus and species *Aceticella autotrophica* is proposed. Genomic analysis identified genes encoding all enzymes necessary for autotrophic acetogenesis. The oxidative module was represented by bifurcation hydrogenase, the reductive module, by the Wood-Ljungdahl pathway (WLP), and the energy module, by ECH hydrogenase and H+-dependent ATP synthase. The proposed metabolic patterns of *Ac. autotrophica* 3443-3Ac^T during growth on H_2/CO_2, HCOOH and CO suggest the operation of the same enzymatic apparatus; the differences lie in the involvement of monofunctional CO dehydrogenase during growth on CO and the opposite direction of HDCR operation during growth on formate. For *Ac. autotrophica* 3443-3Ac^T, a new energy conservation mechanism was proposed – "facilitated" acetogenesis, in which the catabolic function of WLP, associated with the reoxidation of reducing equivalents, is in part taken over by a NAD(P)H-dependent polysulfide reductase. The increase in the amount of oxidized $NAD(P)^+$ allows bifurcation hydrogenase to more efficiently produce low-potential ferredoxins necessary for the generation of membrane potential by the ECH complex.

Strain 3819-GS1 grew exclusively under chemolithoautotrophic conditions with H_2 as the electron donor and CO_2 as the electron acceptor and carbon source, and the only metabolic product was acetate. Thus, it was established that strain 3819-GS1 is also an obligate autotrophic acetogenic bacterium. New isolate is not capable of reducing external electron acceptors. Based on phylogenomic studies, a new genus and species, *Acetitalea autotrophica*, was proposed. Genome analysis revealed significant differences in the metabolism of the new isolate from the metabolism of all currently known acetogenic microorganisms. It was suggested that the oxidative module is represented by bifurcation hydrogenase, the reductive module, by the reductive glycine pathway, and the energy module, by an 11-subunit Nuo-like membrane complex and H+-dependent ATP synthase. Thus, we have proposed a new mechanism of "glycine" acetogenesis.

Criteria have been developed for classifying microorganisms as potential acetogens ("classical" and hypothetical), as well as criteria for assessing the presence or absence of energy conservation mechanisms known for acetogens (Rnf-complex and ECH-hydrogenase). 280 potential "classical" acetogens were identified in 18 phyla. The majority belonged to the phyla Bacillota_A (178) and Bacillota_B-G (22 to 4) and were scattered across the taxonomic (phylogenetic) diversity of this group, without being

numerically predominant in any of its lineages. These results are consistent with the idea of vertical inheritance of the capacity for “classical” acetogenesis and with our conclusion about the vertical inheritance of this capacity, made on the basis of phylogenetic trees of key determinants. 61 genomes of potential unusual acetogens (WLP in the absence of Rnf and ECH) were discovered. However, in cases where these were genomes of cultivated microorganisms, they most often belonged to sulfate reducers from the orders *Desulfotomaculales* and *Desulfitobacteriales*, and their WLP most likely performs a function other than acetogenesis. 187 genomes of potential “glycine” acetogens were discovered, and only six of them lacked the ECH and Rnf genes. Thus, the pattern of determinants observed in our “glycine” acetogen *At. autotrophica* 3819-GS1 is quite rare. Interestingly, of these 187 potential acetogens, 27 had determinants of both glycine pathway and WLP.

Conclusion: Thus, the first obligate chemolithoautotrophic acetogenic bacteria, *Aceticella autotrophica* and *Acetitalea autotrophica*, were isolated and characterized. For *Ac. autotrophica* 3443-3Ac^T, a new energy conservation mechanism was proposed – “facilitated” acetogenesis, in which the catabolic function of WLP, associated with the reoxidation of reducing equivalents, is in part taken over by a NAD(P)H-dependent poltsilfide reductase. For *At. autotrophica* 3819-GS1^T, a new mechanism of “glycine” acetogenesis was proposed. Analysis of genomic databases revealed a widespread distribution of novel energy conservation mechanisms among bacteria.

Funding: The study is supported by the grant from the Russian Science Foundation No. 24-14-00177.

doi 10.18699/bgrs2024-8.2-09

Изменение микробиоты кишечника у крыс-самок Wistar при индукции экспериментального рака молочной железы

Черкас В.[1], Кабаков А.[1]*, Повещенко А.[1], Казаков О.[1], Леляк А.[2], Бодрова Н.[3]

[1] *Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия*

[2] *ООО НПФ «Исследовательский центр», р.п. Кольцово, Новосибирская область, Новосибирск, Россия*

[3] *ФГБОУ ВО Новосибирский государственный аграрный университет, Новосибирск, Россия*

* *valeriya_korol@mail.ru*

Ключевые слова: рак молочной железы (РМЖ); кишечная микробиота; индукция; N-метил-N-нитрозомочевина (МНМ); дисбиоз

Мотивация и цель: Учитывая современные знания о роли микробиоты в развитии канцерогенеза, актуальным является исследование микробиоты при РМЖ, так как рак молочной железы является одним из самых распространенных раков во всем мире [1]. Микробиоту кишечника можно рассматривать как потенциальный диагностический онкобиомаркер с одной стороны и патогенетический фактор, на который будет направлена терапия и профилактика РМЖ, с другой стороны.
Целью исследования стал анализ изменений количественного и качественного состава кишечной микробиоты у крыс линии Wistar при химической индукции рака молочной железы.
Методы и алгоритмы: Изучение параметров микробной колонизации толстой кишки проводили у самок крыс линии Wistar (n = 40) в возрасте 3 месяцев, массой 210–250 г. Индуцирование РМЖ выполнялось 5-кратным подкожным введением N-метил-N-нитрозомочевины (Sigma) с интервалом в 7 дней в область 2-й правой молочной железы. Животные, участвующие в эксперименте, были разделены на 2 группы: 1-я: интактные животные (n = 20); 2-я: группа крыс, которым вводили МНМ для индукции РМЖ (n = 20). У интактных крыс (n = 20) забор кала для исследования микрофлоры толстого кишечника выполняли в 1-й, 14-й и 35-й дни параллельно с группой индукции РМЖ. У крыс 2-й группы (n = 20) забор материала для бактериологического исследования проводили в 1-й день (до инъекции N-метил-N-нитрозомочевины), на 14-й и 35-й дни индукции РМЖ. Материалом для бактериологического исследования служили фекалии исследуемых животных. Пробу для исследования брали из последней порции фекалий. Показатели для оценки особенностей микробиоты кишечника для испытаний были выбраны согласно рекомендациям Отраслевого стандарта 91500.11.0004-2003 «Протокол ведения больных. Дисбактериоз кишечника» [2]. Для культивирования бактерий использовали различные питательные среды в соответствии с инструкциями производителя. Результаты принимались в расчет по числу выросших колоний с определением культуральных, морфологических (микроскопия) и тинкториальных свойств (окраска по Граму) после истечения сроков инкубации. Количество выделенных микроорганизмов рассчитывалось согласно формуле: $КОЕ/г = К \times 10 \times n$, где $К$ – количество выросших колоний; n – разведение суспензии; 10 – коэффициент пересчета на 1 см^3 суспензии при посеве 0.1 см^3 (0.1 см^3 составляет 1/10 см^3). Полученный результат переводился в десятичный

логарифм числа колониеобразующих единиц в 1 г исследуемого материала (log_{10} КОЕ/г). Статистическую обработку данных проводили с использованием пакета программ «Statistica 10.0». Достоверность различия рассчитывали по U-критерию Манна–Уитни и принимали при значениях $p < 0,05$.

Результаты: В фекалиях, полученных от интактных крыс, не наблюдали изменений среди типичных представителей микробиоты желудочно-кишечного тракта на 1-й, 14-й и 35-й дни исследований. Средняя концентрация *Escherichia coli* (типичных) составила 25.8×10^5 КОЕ/г (6.4±0.2 log10 КОЕ/г), при этом уровень *Escherichia coli* (лактозонегативных) был $<10^5$ КОЕ/г. Уровень лакто-бифидобактерий, клостридий и дрожжеподобных грибов рода Candida находился в допустимых пределах: от $< 10^4$ КОЕ/г для грибов рода Candida до 10^9 КОЕ/г для бифидобактерий. Стоит отметить, что *Staphylococcus aureus* и *Staphylococcus saprophyticus* не высеивались. Микрофлора толстого отдела кишечника во 2-й группе крыс в 1-й день (до инъекции N-метил-N-нитрозомочевины) достоверно не отличалась от показателей интактных животных. К 14-ому дню индукции РМЖ в полученных образцах кала отмечено увеличение концентрации типичных *Escherichia coli*, которое составило 426.4×10^5 КОЕ/г (7.6±0.8 log10 КОЕ/г), что в 16.5 раза больше, чем в группе интактных крыс. На 35-й день индукции опухоли выявлено достоверное появление патогенной микрофлоры, в частности 60 % случаев *Staphylococcus aureus* в средней концентрации, составившей 9.4×10^5 КОЕ/г (6.0±0.2 log10 КОЕ/г). Было обнаружено наличие *Staphylococcus saprophyticus* в 100 % случаев на уровне $29x10^5$ КОЕ/г (6.5±0.2 log10 КОЕ/г). Отмечено сохранение типичной *Escherichia coli* на высоком уровне 125.6×10^5 КОЕ/г (7,1±0,3 log10 КОЕ/г) и возрастание лактозонегативной *Escherichia coli* до значений 19.3×10^5 КОЕ/г (6.2±0.6 log10 КОЕ/г), что в 1.5 раза выше по сравнению с показателями в группе интактных крыс и у крыс на 14-й день индукции РМЖ. У нескольких крыс на 35-й день индукции РМЖ обнаружено наличие в фекалиях условно-патогенной энтеробактерии *Proteus vulgaris* в значимой концентрации 10^6 КОЕ/г.

Выводы: Таким образом, установлено, что выделенные бактерии принадлежали к 3 типам, 4 классам, 5 порядкам, 6 семействам, 6 родам домена бактерий. Также было выделено 2 рода грибов, относящихся к порядку *Saccharomycetales*. Выделенные у крыс первой и второй групп бактерии в большинстве своем принадлежали к типам *Actinobacteria*, *Firmicutes* и *Proteobacteria*, являющимися типичными представителями микробиоты желудочно-кишечного тракта крыс, близкой по своему составу к микробиоте кишечника человека [3]. При этом количественный и качественный состав кишечной микробиоты у крыс при индуцировании РМЖ химическим канцерогеном (МНМ) изменялся достоверно с возрастанием концентрации патогенных микроорганизмов к 35-му дню индукции РМЖ, что свидетельствует о формирующемся дисбиозе кишечной микробиоты, который по ряду исследований является предрасполагающим фактором в возникновении хронического воспаления – основополагающего звена, способствующего развитию опухолей в различных органах, включая молочную железу [4].

Changes in the intestinal microbiota in female Wistar rats during the induction of experimental mammary cancer

Cherkas V.[1]*, Kabakov A.[1], Poveshchenko A.[1], Kazakov O.[1], Lelyak A.[2], Bodrova N.[3]

[1] *Recearch Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia*

[2] *NPF "Research Center", Novosibirsk region, Koltsovo, Russia*

[3] *Novosibirsk State Agrarian University (NSAU), Novosibirsk, Russia*

** valeriya_korol@mail.ru*

Key words: breast cancer (BC); intestinal microbiota; induction; N-methyl-N-nitrosourea (MNM); dysbiosis

Motivation and Aim: Considering modern knowledge about the role of microbiota in the development of carcinogenesis, the study of microbiota in breast cancer is relevant, since breast cancer is one of the most common cancers worldwide [1]. Intestinal microbiota can be considered as a potential diagnostic oncobiomarker, on the one hand, and a pathogenetic factor, which will be targeted in the treatment and prevention of breast cancer, on the other hand. The aim of the study was to analyze changes in the quantitative and qualitative composition of the intestinal microbiota in Wistar rats during chemical induction of breast cancer.

Methods and Algorithms: The study of the parameters of microbial colonization of the colon was carried out in female Wistar rats ($n = 40$) aged 3 months, weighing 210–250 g. Induction of breast cancer was carried out by 5-fold subcutaneous injection of N-methyl-N-nitrosourea (Sigma) with an interval of 7 days in the area of the 2nd right breast. The animals participating in the experiment were divided into 2 groups: 1st: intact animals ($n = 20$); 2nd: group of rats that were administered MNM to induce breast cancer ($n = 20$). In intact rats ($n = 20$), feces were collected to study the microflora of the large intestine on 1st, 14th and 35th days in parallel with the breast cancer induction group. In rats of the 2nd group ($n = 20$), material was collected for bacteriological examination on the 1st day (before the injection of N-methyl-N-nitrosourea), on the 14th and 35th days of breast cancer induction. The material for bacteriological research was the feces of the studied animals. The sample for the study was taken from the last portion of feces. Indicators for assessing the characteristics of the intestinal microbiota for testing were selected according to the recommendations of Industry Standard 91500.11.0004-2003 "Protocol for the management of patients. Intestinal dysbiosis" [2]. Various culture media were used to cultivate bacteria according to the manufacturer's instructions. The results were taken into account by the number of grown colonies with the determination of cultural, morphological (microscopy) and tinctorial properties (Gram stain) after the incubation period had expired. The number of isolated microorganisms was calculated according to the formula: $CFU/g = K \times 10 \times n$, where K is the number of grown colonies; n – suspension dilution; 10 is the conversion factor per 1 cm^3 of suspension when sowing 0.1 cm^3 (0.1 cm^3 is 1/10 cm^3). The obtained result was converted into the decimal logarithm of the number of colony-forming units in 1 g of the test material (log10 CFU/g). Statistical data processing was carried out using the software package "Statistica 10.0". The significance of the difference was calculated using the Mann–Whitney U test and accepted at p values < 0.05.

Results: In feces obtained from intact rats, no changes were observed among typical representatives of the gastrointestinal tract microbiota on 1st, 14th and 35th days of the

study. The average concentration of *Escherichia coli* (*typical*) was 25.8×10^5 CFU/g (6.4 ± 0.2 log10 CFU/g), while the level of *Escherichia coli* (*lactose-negative*) was $< 10^5$ CFU/g. The level of lactobifidobacteria, clostridia and yeast-like fungi of the genus Candida was within acceptable limits: from $< 10^4$ CFU/g for fungi of the genus Candida to 10^9 CFU/g for bifidobacteria. It is worth noting that *Staphylococcus aureus* and *Staphylococcus saprophyticus* were not cultured. The microflora of the large intestine in the 2nd group of rats on the 1st day (before the injection of N-methyl-N-nitrosourea) did not differ significantly from the indicators of intact animals. By the 14th day of breast cancer induction, an increase in the concentration of *typical Escherichia coli* was noted in the obtained stool samples, which amounted to 426.4×10^5 CFU/g (7.6 ± 0.8 log10 CFU/g), which is 16.5 times more than in group of intact rats. On the 35th day of tumor induction, a significant appearance of pathogenic microflora was revealed, in particular 60 % of cases of *Staphylococcus aureus* in an average concentration of 9.4×10^5 CFU/g (6.0 ± 0.2 log10 CFU/g). The presence of *Staphylococcus saprophyticus* was detected in 100 % of cases at a level of 29×10^5 CFU/g (6.5 ± 0.2 log10 CFU/g). It was noted that *typical Escherichia coli* remained at a high level of 125.6×10^5 CFU/g (7.1 ± 0.3 log10 CFU/g) and an increase in *lactose-negative Escherichia coli* to values of 19.3×10^5 CFU/g (6.2 ± 0.6 log10 CFU/g), which is 1.5 times higher compared to the indicators in the group of intact rats and in rats on the 14th day of breast cancer induction. In several rats on the 35th day of breast cancer induction, the presence in the feces of the opportunistic enterobacterium *Proteus vulgaris* was detected at a significant concentration of 10^6 CFU/g.

Conclusion: Thus, it was established that the isolated bacteria belonged to 3 types, 4 classes, 5 orders, 6 families, 6 genera of the bacterial domain. Also, 2 genera of fungi belonging to the order *Saccharomycetales* were isolated. The bacteria isolated from rats of the first and second groups mostly belonged to the types *Actinobacteria*, *Firmicutes* and *Proteobacteria*, which are typical representatives of the microbiota of the gastrointestinal tract of rats, similar in composition to the human intestinal microbiota [3]. At the same time, the quantitative and qualitative composition of the intestinal microbiota in rats when breast cancer was induced by a chemical carcinogen (MNM) changed significantly with an increase in the concentration of pathogenic microorganisms by the 35th day of breast cancer induction, which indicates the emerging dysbiosis of the intestinal microbiota, which, according to a number of studies, is a predisposing factor in the occurrence chronic inflammation is a fundamental link that promotes the development of tumors in various organs, including the mammary gland [4].

Список литературы/References

1 Sung H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi 10.3322/caac.21660

2 ОСТ 91500.11.0004-2003. Отраслевой стандарт. Протокол ведения больных. Дисбактериоз кишечника Режим доступа: https://docs.cntd.ru/document/1200119089
[OST 91500.11.0004-2003. Industry standard. Patient management protocol. Intestinal dysbiosis Available at: https://docs.cntd.ru/document/1200119089]

3 Flemer B. et al. Fecal microbiota variation across the lifespan of the healthy laboratory rat. *Gut Microbes*. 2017;8(5):428-439. doi 10.1080/19490976.2017.1334033

4 Buchta Rosean C. et al. Preexisting commensal dysbiosis is a host-intrinsic regulator of tissue inflammation and tumor cell dissemination in hormone receptor-positive breast cancer. *Cancer Res.* 2019;79(14):3662-3675. doi 10.1158/0008-5472.CAN-18-3464

doi 10.18699/bgrs2024-8.2-10

Bioleaching of copper-zinc concentrates under different conditions

Bulaev A.*, Artykova A., Elkina Y., Kolosov A., Nechaeva A., Beletsky A., Kadnikov V., Melamud V., Mardanov A.

Research Centre of Biotechnology RAS, Moscow, Russia

** bulaev.inmi@yandex.ru*

Key words: bioleaching; polymetallic sulfide concentrates; copper; zinc; acidophilic microorganisms; metabarcoding

Motivation and Aim: Substandard copper concentrates do not meet quality standards for pyrometallurgical processing due to the high content of impurities such as zinc, arsenic, and other metals and metalloids. For the processing of this type of raw material, hydrometallurgical technologies are promising, including biohydrometallurgical technologies, which make it possible to process raw materials with a low content of target components and a high content of harmful impurities. Substandard copper concentrates include concentrates with high arsenic content in copper minerals such as tennantite and enargite. Currently, the possibility of processing concentrates with a high content of tennantite and enargite using various hydrometallurgical technologies (autoclave leaching, biohydrometallurgy, etc.) is actively studied. An analysis of works on the bioleaching of tennantite and enragite minerals, as well as concentrates containing these minerals, showed that bioleaching is a promising method for processing such concentrates, however, the information available in the literature on the dependence of the rate of bioleaching of arsenic-containing copper minerals on various factors is incomplete and even contradictory. Therefore, the purpose of this work was to study the process of bioleaching of arsenic-containing polymetallic concentrates containing tennantite, chalcopyrite and sphalerite under different conditions. The dependence of non-ferrous metal leaching on process temperature (in the range of 40–60 °C) and the use of CO_2 and molasses bioreactors as carbon sources for the microbial population was studied. Previously, our works showed that various carbon sources can affect the biooxidation of pyrite-arsenopyrite concentrates at elevated temperatures [1]. Therefore, in this work, we investigated the influence of carbon sources on the bioleaching of copper concentrates, since they can also increase the rates of bioleaching of copper.
Methods and Algorithms: For experiments, sulfide concentrates, the main minerals of which were chalcopyrite ($CuFeS_2$), tennantite ($Cu_{12}As_4S_{13}$), sphalerite (ZnS), pyrite (FeS_2), and quartz (SiO_2) were used. The first concentrate contained 6.2 % copper, 7.3 % zinc, 1.7 % arsenic, 24.4 % iron, 31.9 % sulfide sulfur, the second concentrate contained 16.0 % copper, 5.3 % zinc, 1.7 % arsenic, 28.0 % iron, 33.2 % sulfide sulfur. Thus, in the first concentrate a large proportion of copper was contained in tennantite, while in the second concentrate the main part of copper was contained in chalcopyrite. Bioleaching was carried out in laboratory scale reactors in a continuous The residence time was 10 days, the pulp density was 1 : 10 (S : L). For the concentrate 1, bioleaching was performed at 40, 45, 50, 55, and 60 °C. For the concentrate 2, bioleaching was performed at 45 and 55 °C. The influence of carbon sources on the biooxidation process

was investigated by adding 0.02 % molasses to the pulp or by supplying CO_2. Microbial populations were analyzed by metabarcoding of V3-V4 fragments of 16S rRNA genes.
Results: Bioleaching of concentrate 1 in the temperature range of 40–60 °C showed that the highest rate of copper recovery was achieved at a temperature of 45–50 °C (23–27 %), and increasing the temperature to 55 and 60 °C led to a decrease in copper leaching to 20–22 and 10–20 %, respectively. In the temperature range of 45–50 °C, the degree of copper leaching was weakly dependent on the use of additional carbon sources, while at 60 °C, leaching increased when using carbon dioxide (up to 20 %), while in variant with molasses and in the control experiment (without additional source carbon), copper leaching reached about 10 %.
When bioleaching concentrate 2, increasing the temperature from 45 to 55 °C led to an increase in the degree of copper leaching from 25–28 to 36–52 %. At the same time, the degree of leaching at 45°C depended to a small extent on the carbon source, while at 55 °C the highest degree of copper leaching was when molasses was used.
The composition of microbial populations differed during bioleaching of concentrates 1 and 2, as well as when process conditions changed. During the bioleaching of concentrate 1 at lower temperatures, archaea of the genus *Ferroplasma* and bacteria of the genera *Acidithiobacillus* and *Sulfobacillus* predominated. An increase in temperature led to the predominance of archaea of the genus *Acidiplasma*. During bioleaching of the concentrate 2, archaea of the genus *Ferroplasma*, and bacteria of the genera *Acidithiobacillus* and *Sulfobacillus* were predominant. Bacteria of the genus *Sulfobacillus* in this case were dominant at a temperature of 55 °C.
Conclusion: Thus, in this work, when studying the process of bioleaching of arsenic-containing polymetallic concentrates containing chalcopyrite, tennantite and sphalerite, at different temperatures and using different carbon sources, it was shown that the use of different carbon sources and an increase in temperature makes it possible to intensify the process of copper bioleaching. Moreover, a comparison of the results obtained with the results obtained in previous works [2, 3] shows that such effects can be determined by the ratio of sulfide minerals in the concentrate. The observed effect of the studied factors on the bioleaching process is largely determined by their impact on the microbial population that carries out the bioleaching process, since changes in the composition of the microbial populations of bioreactors were observed when the process conditions changed.
Funding: The study is supported by the Russian Science Foundation (grant No. 21-64-00019).

References

1. Bulaev A., Kadnikov V., Elkina Y., Beletsky A., Melamud V., Ravin N., Mardanov A. Shift of microbial population of bioleach reactors determined by carbon sources and temperature. *Biology*. 2023;12:1411. doi 10.3390/biology12111411
2. Elkina Y.A., Bulaev A.G., Melnikova E.A., Melamud V.S. Bioleaching of enargite and tennantite by moderately thermophilic acidophilic microorganisms. *Microbiology*. 2020;89:413-424. doi 10.1134/S0026261720040050
3. Elkina Y.A., Melamud V.S., Bulaev A.G. Bioleaching of a copper-zinc concentrate with high arsenic content. *Microbiology*. 2021;90:78-86. doi 10.1134/S002626172006003X

doi 10.18699/bgrs2024-8.2-11

Approaches to the analysis of fungal communities using the example of the human mycobiome patient with COVID-19

Krivonos D.[1, 2]*, Fedorov D.[1], Orlov A.[1, 3], Klimina K.[4], Veselovsky V.[4], Kovalchuk S.[1], Pavlenko A.[1], Ilina E.[1]

[1] *Research Institute for Systems Biology and Medicine (RISBM), Moscow, Russia*

[2] *Department of Molecular and Translational Medicine, Moscow Institute of Physics and Technology, State University, Dolgoprudny, Russia*

[3] *I.M. Sechenov First Moscow State Medical University, Moscow, Russia*

[4] *Federal Research and Clinical Centre of Physical and Chemical Medicine, Federal Medical and Biological Agency of Russia, Russia*

* *danil01060106@gmail.com*

Key words: fungal community; metagenomic; microbiology

Motivation and Aim: The human microflora is a complex composite system that includes various microorganisms. To an extent, the methods of modern biology are based on the study of the major component of the human microflora, represented by the bacterial part. However, there is a growing interest in the study of fungal microflora in the scientific community. Interest in this topic has grown against the COVID-19 pandemic. Thus, in people with COVID-19, an increased presence and change in the spectrum of fungal microorganisms was described [1]. At the same time, complications from fungal infections have become the cause of severe disease [2, 3]. At the same time, there is no understanding of the patterns of formation of a complicated picture of the course of COVID-19 through the addition of nosocomial bacterial or fungal flora.

Today, there are a number of studies showing that COVID-19 affects various human organs, including the intestines and its microflora. A significant amount of information has already accumulated in the literature on the study of changes in bacterial flora. However, much less research has focused on the human fungal microflora (i. e., mycobiome).

The aim of the work was to assess changes in the human mycobiome against the background of COVID-19, as well as to identify factors associated with changes in the composition of the fungal community. In our study, we carried out a taxonomic annotation of the mycobiome of patients with COVID-19 at the time of hospitalization and compared it with the state of the fungal microflora formed during treatment for coronavirus infection. The work also demonstrates differences between groups of patients at the first and second time points, which allows us to obtain a basic understanding of the trend in the direction of changes in the mycobiome during COVID-19 therapy.

The fungal part of the human microflora represents a minor component, thereby limiting the range of methods for studying it. Currently, the most common technology for studying fungal communities is amplicon sequencing of the internal transcribed spacer (ITS). However, in practice, only ITS1 or ITS2 regions from the full ITS cluster are used. In turn, they make it possible to reflect the taxonomy of different fungi with varying efficiency [4]. In turn, the full-length region ITS1-5.8S-ITS2 has a much stronger

phylogenetic signal and will better describe the taxonomy and composition of the fungal community.

Another rather important difference between the study of fungal ITS and bacterial 16S is the relatively high variability in the lengths of ITS amplicon sequences. Polymerase chain reaction, which is used in the preparation of libraries for sequencing, in turn, can have a certain bias in the assessment of fungal composition. Based on this, one of the tasks of this work is to find an optimal procedure for normalizing ITS to the length of the sequence itself.

Methods and Algorithms: The study included fecal samples (n = 117) obtained from patients during the first day of hospitalization (first time point) and closer to recovery (second time point). The control group consisted of 8 fecal samples taken from healthy donors. With this study, we would also like to expand the geographical understanding of the fungal microflora of the intestine. For the assembled collection of samples, a dataset of sequencing full-length sections of the ITS1-5.8-ITS2 regions was obtained on the Illumina MySeq platform with a library size of 300 + 300.

In the resulting data set, some of the sequences, as expected, do not cover the middle region of 5.8S rRNA and for this reason, it is not possible to use standard approaches for processing amplicon data using dada2. In our work, we applied sequence mapping to the fungal ITS UNITE database using bwa. Then the counts were calculated for each covered sequence in the database. We considered a successful hit to be those records that cover more than 95 % of the maximum possible length that we could cover with the available library size, in our case 600 bp.

We decided to further normalize the resulting tables with record counts in UNITE to the lengths of aligned records by calculating the TPM metric, which is actively used in transcriptomics. In addition, we also analyzed the clr transformed shares of the relative representations of the counts obtained on the raw matrix.

Agglomeration of entries in the table was carried out using the tip_glom function on a phylogenetic tree built using mafft + fasttree for sequences from the UNITE reference database. Clustering of samples by enterotype was assessed using DirichletMultinomial. DESeq2 was used to identify differentially represented fungi. Reduced dimensionality imaging was performed using PCoA. To identify factors associated with mycobiome diversity, we used the nonparametric PERMANOVA test.

Results: The major part in the analyzed array is represented by fungi of the genus *Candida*, in particular *Candida albicans*, which was found in all samples. However, the minor components of the mycobiome may vary. Thus, in a number of samples, the presence of representatives of the non-albigans *Candida sp.* group is noted: *Candida parapsilosis, Candida sake, Candida dubliniensis, Candida tropicalis* and *Candida mesenterica*. A significant part of the samples also includes representatives of the *Dipodascacea* family - *Yarrowia lipolytica* and *Geotrichum candidum*. The analysis also revealed that *Geotrichum candidum* is the main driver of differences between time points; its completion is characteristic of the first time point. *Yarrowia lipolytica* and *Saccharomyces cerevisiae* tend to be overestimated at the first point and underestimated at the second.

There were no obvious major differences in the taxonomic profile of fungal compositions between the different groups of people studied. However, as a result of the study, it was noted that the fungal alpha diversity of the group of sick patients was generally higher than that of healthy patients (p-value < 0.05), but no statistically significant differences were found between the time points themselves. However, it is

worth noting that alpha diversity in patients at the first time point is slightly higher than at the second, which indirectly hints at a trend towards the return of fungal microflora to a healthy state.

In the analysis of factors of individual characteristics of patients and characteristics of the treatment procedure using PERMANOVA 2, it was determined to be statistically significant. These are the presence of inflammatory bowel disease (p-value < 0.05) and the presence of diabetes (p-value < 0.05). The literature also shows an association of fungal microflora with inflammatory bowel disease and diabetes [5, 6]. At the same time, it was noted that the total percentage of explained variance in beta diversity exceeds 60 %, which suggests that the diversity of human fungal microflora is largely explained by individual characteristics of the person and the treatment procedure.

For a more accurate representation and association of the mycobiome with various factors, further research is needed that will expand the array of available information

Funding: The study is supported by state task No. 122030900064-9 of Rospotrebnadzor.

References

1. Zuo T., Zhan H., Zhang F., Liu Q., Tso E.Y.K., Lui G.C.Y., Chen N., Li A., Lu W., Chan F.K.L., Chan P.K.S., Ng S.C. Alterations in fecal fungal microbiome of patients with COVID-19 during time of hospitalization until discharge. *Gastroenterology*. 2020;159(4):1302-1310. doi 10.1053/j.gastro.2020.06.048
2. Negm E.M., Mohamed M.S., Rabie R.A., Fouad W.S., Beniamen A., Mosallem A., Tawfik A.E., Salama H.M. Fungal infection profile in critically ill COVID-19 patients: a prospective study at a large teaching hospital in a middle-income country. *BMC Infect Dis*. 2023;23(1):246. doi 10.1186/s12879-023-08226-8
3. Hoenigl M., Seidel D., Sprute R., Cunha C., Oliverio M., Goldman G.H., Ibrahim A.S., Carvalho A. COVID-19-associated fungal infections. *Nat Microbiol*. 2022;7(8):1127-1140. doi 10.1038/s41564-022-01172-2
4. Mbareche H., Veillette M., Bilodeau G., Duchaine C. Comparison of the performance of ITS1 and ITS2 as barcodes in amplicon-based sequencing of bioaerosols. *PeerJ*. 2020;8:e8523. Doi 10.7717/peerj.8523
5. Hsu C., Ghannoum M., Cominelli F., Martino L.D. Mycobiome and inflammatory bowel disease: role in disease pathogenesis, current approaches and novel nutritional-based therapies. *Inflamm Bowel Dis*. 2023;29(3):470-479. doi 10.1093/ibd/izac156
6. Wang L., Zhang K., Zeng Y., Luo Y., Peng J., Zhang J., Kuang T., Fan G. Gut mycobiome and metabolic diseases: The known, the unknown, and the future. *Pharmacol Res*. 2023;193:106807. doi 10.1016/j.phrs.2023.106807

doi 10.18699/bgrs2024-8.2-12

Microbiome manipulation of potato with biostimulator *Rhodococcus qingshengii* VKM Ac-2784D

Petrushin I.S.*, Markova Yu.A., Morits A.S., Gutnik D.I., Filinova N.V.
Siberian Institute of Plant Physiology and Biochemistry, SB RAS, Irkutsk, Russia
** ivan.kiel@gmail.com*

Key words: microbiome; endosphere; rhizosphere; potato; *R. qingshengii* VKM Ac-2784D; genomics

Motivation and Aim: Plants' microbiome helps them to adapt to the environment conditions and keep the organism healthy. Microbiome consists of vast range of microorganisms, which form the trophic chain, keeping stable when environment conditions change. Microbiome changing can infer into trophic chains and impact negatively to the health of the plant host. Bacteria of genus Rhodococcus are typical rhizosphere inhabitants. Like other actinomycetes they synthesize a broad range of bioactive compounds: biosurfactants, phytohormones, siderophores, antibiotics. In our previous studies we isolated the strain *Rhodococcus qingshengii* VKM Ac-2784D, which can digest the alkanes and surface-active substances [1, 2] and it has lactonase activity, which infers the quorum sensing (also called quorum quenching) [3]. We suggest that the strain can be used as the fertilizer agent, it will impact on the rhizosphere and endosphere microbiome. The complete genome of *Rhodococcus qingshengii* VKM Ac-2784D is available NCBI GenBank [4]. Genome analysis with antiSMASH revealed a number of biosynthetic gene clusters of siderophores and some antibiotics: monensin, rifamorpholine, thiolutin, acyldepsipeptide 1, tetronasin, hygromycin A, kirromycin, diisonitrile antibiotic SF2768. To evaluate the diversity shift of rhizosphere and endosphere communities we performed the amplicon sequencing study.
Methods and Algorithms: We used the “Lugovskoy” variety in this study as it common for the region. Tubers were treated for 24 h by bacterial suspension of *Rhodococcus qingshengii* VKM Ac-2784D. The secondary treatment was performed on plant flowering period (21 July 2023). Sample collection was performed at three stages: seedling, plant flowering and potato harvesting. DNA was isolated from rhizosphere soil, leaves and tubers samples (38 samples in total). PCR product was further sequenced on BGIseq platform at shared access center “Genomica” (Novosibirsk). Taxonomic profiling was performed using USEARCH/SINTAX algorithm (RDP 16S database).
Results: After treatment the mean number of tubers per plant increased by 70 % and total yield increased by 55 %, however the mean size (44 × 30 mm) and mass (33–36 g) of tubers stayed almost the same. Microbiome analysis showed that representatives of phylum Proteobacteria, Acidobacteria, Bacteroidetes, Actinobacteria and Firmicutes predominated in the soil samples. Preliminary data indicate that treatment with *Rhodococcus* cell suspension led to a decrease in the

abundance of Proteobacteria and Bacteroidetes. In contrast, the content of Acidobacteria, Firmicutes, and Verrucomicrobia increased (Fig. 1). These results have common features with other inoculation practices, that are described in the literature [5], but other factors – soil type, irrigation regime, pathogens influence has to be evaluated.

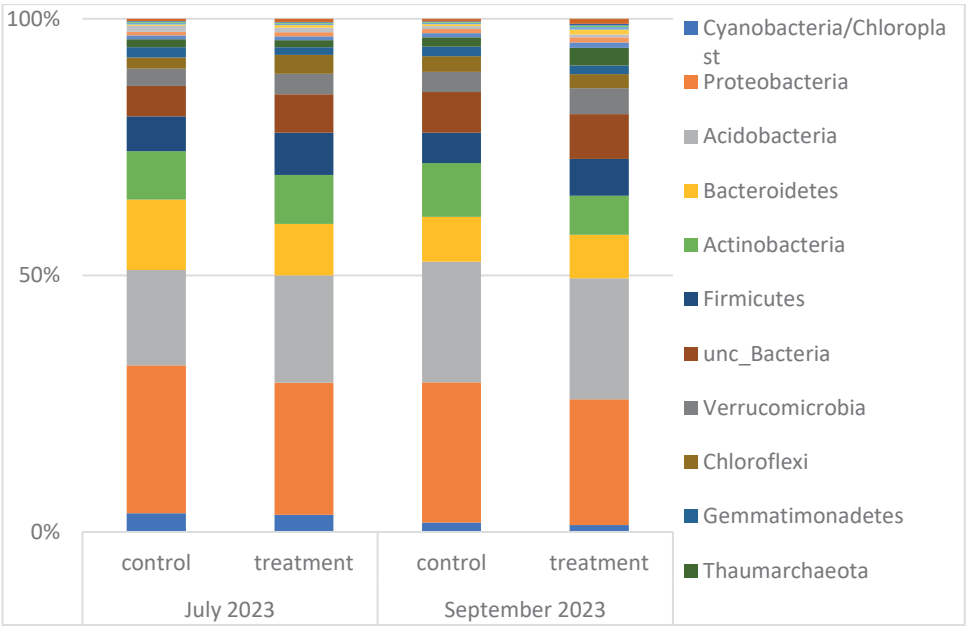


Fig. 1. Potato rhizosphere microbiome taxonomic structure for control and treatment samples; sampling was performed two times: at plant flowering and harvesting

Conclusion: As a result of our study, we got the knowledge of influence of *Rhodococcus qingshengii* VKM Ac-2784D after using it as the fertilizer agent. The samples from potato rhizosphere and endosphere could be used as the source of microorganisms' communities to form the collections for further study.

Funding: This work was supported by the Russian Science Foundation (RSF; project No. 23-26-10049).

References

1. Belovezhets L.A. et al. Possible pathways for destruction of polyaromatic hydrocarbons by some oil-degrading bacteria isolated from plant endosphere and rhizosphere. Appl Biochem Microbiol. 2017;53:68-72. doi 10.1134/S0003683817010069
2. Belovezhets L.A. et al. Use of rhizosphere microorganisms for bioremediation of oil contaminated soils. *IOP Conf. Ser.: Earth Environ. Sci.* 2020;408:012086. doi 10.1088/1755-1315/408/1/012086
3. Tretyakova M.S. et al. Possible use of oil-degrading microorganisms for protection of plants growing under conditions of oil pollution. *IOP Conf. Ser.: Earth Environ. Sci.* 2019;315:072046 doi 10.1088/1755-1315/315/7/072046
4. Petrushin I. S. et al. Complete Genome Sequence of Rhodococcus qingshengii Strain VKM Ac-2784D, Isolated from Elytrigia repens Rhizosphere. *Microbiol. Res. Announc.* 2021;10(11). doi 10.1128/MRA.00107-21
5. Petrushin I.S. et al. Potato Microbiome: Relationship with Environmental Factors and Approaches for Microbiome Modulation. *Int. J. Mol. Sci*. 2024;25:750. doi 10.3390/ijms25020750

doi 10.18699/bgrs2024-8.2-13

Hi-C metagenomic study of antimicrobial resistance genes spread in critically ill patients

Revel-Muroz A.Z.[1]*, Chistyakov A.S.[2], Sonets I.V.[2], Ivanova V.A.[2], Vasiluev P.A.[2, 3], Surovoy Y.A.[4, 5, 6], Kozlovskaya L.I.[7, 8], Fursova N.K.[10], Ulianov S.V.[2, 9], Tyakht A.V.[1, 2]

[1] *Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology Russian Academy of Sciences, Moscow, Russia*

[2] *Institute of Gene Biology Russian Academy of Sciences, Moscow, Russia*

[3] *Research Center for Medical Genetics, Moscow, Russia*

[4] *Mechnikov Research Institute for Vaccines and Sera, Moscow, Russia*

[5] *Faculty of Medicine, M.V. Lomonosov Moscow State University, Moscow, Russia*

[6] *Clinical Hospital #1 MEDSI, Otradnoe, Moscow Oblast, Russia*

[7] *FSASI "Chumakov FSC R&D IBP RAS" (Institute of Poliomyelitis), Moscow, Russia*

[8] *Sechenov First Moscow State Medical University, Moscow, Russia*

[9] *Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia*

[10] *State Research Center for Applied Microbiology and Biotechnology, Obolensk*

* *revelanastasia@gmail.com*

Key words: Hi-C metagenomics; antimicrobial resistance

Motivation and Aim: Antimicrobial resistance (AMR) is one of the main reasons for the increased mortality of patients in medical institutions. Human microbiome may play a key role in the accumulation and propagation of antibiotic resistance [1]. Therefore, studying mechanisms of AMR genes transmission within bacterial communities will open up new opportunities for improving methods of controlling and limiting their spread in clinical settings. Hi-C metagenomics appears promising in this context, combining information about the genetic and taxonomic composition of a microbiome with the spatial arrangement of genomic fragments [2].

Methods and Algorithms: A total of 12 Hi-C metagenomes of stool samples from patients treated in the ICU were analyzed. In addition, the complete genomes of microorganisms inhabiting sputum were isolated and sequenced (WGS) from 5 of these patients. Metagenome-assembled genomes (MAGs) were reconstructed from the metagenomes. Using information about the strength of Hi-C contacts, we established connections between plasmids and their bacterial hosts. We also conducted a comparative analysis of complete genomes and plasmids found in the lungs and intestines to investigate the evidence of their possible transmission. AMR genes bound to these contigs were also studied.

Results: Combining information on plasmid-to-MAG linkage in gut metagenome for all patients, we built an interaction network of the bacterial genera through plasmids. Final graph included 28 genera from 7 different phyla, showing signs of intra-phylum clusterization. *Firmicutes* and *Proteobacteria* had the highest rate of intra-phylum plasmid transfer, most prevalent genera being opportunistic *Klebsiella*, *Enterococcus*, *Clostridia* and *Streptococcus*, as well as commensal *Limosilactobacillus* and *Lactobacillus*. Also, our data suggest that a significant proportion of plasmids could

potentially cross phylum boundaries. The *Proteobacteria-to-Firmicutes* appeared to be most common among these interactions.

Across the analyzed plasmids, we found a total of 54 unique antibiotic resistance genes belonging to 34 AMR gene families and 21 drug classes. Sixteen genera from 4 phylums carried AMR genes, among them, *Proteobacteria* (composed of *Klebsiella* and *Escherichia*) carried their greatest diversity (37 out of 54). Small fraction of genes were also distributed among the genera related to different phyla.

About 60 of the plasmids were detected in gut metagenome and lung isolates sequences of the same subject, which may indicate their translocation or co-occupation of various niches of the human organism. Interestingly, just one of those events was accompanied by co-colonization with the same bacterial strain of two organs. For a number of other cases, we did not observe species or even genus overlap between the bacterial compositions of two patient's samples. In addition, genes with resistance to beta-lactams, quinolones and sulfonamides antibiotics were associated with intestine-lung intersecting plasmids. Intriguingly, we found same plasmids carrying OXA-1 and CTX-M-15 genes associated with both genomes of *Klebsiella pneumoniae* detected in feces and sputum alongside with *Proteus mirabilis* from lung sample, presenting to us a possible picture of translocation pathways of discussed resistance genes.

Conclusion: Hi-C metagenomics is proving to be a perspective tool in studying the mechanism of plasmid transfer between bacteria. Application of this technology to clinical data will deepen our understanding of the spread of antibiotic resistance genes and its impact on the recovery process of critically ill patients.

Funding: The study is supported by the Russian Science Foundation (No. 19-74-10092).

References

1. Penders J., Stobberingh E.E., Savelkoul P.H., Wolffs P.F. The human microbiome as a reservoir of antimicrobial resistance. *Front Microbiol*. 2013;4:87. doi 10.3389/fmicb.2013.00087
2. Ivanova V. et al. Hi-C Metagenomics in the ICU: exploring clinically relevant features of gut microbiome in chronically critically ill patients. *Front Microbiol*. 2022;12:770323. doi 10.3389/fmicb.2021.770323

doi 10.18699/bgrs2024-8.2-14

Differential expression of two alkane oxidation systems in response to hydrocarbon exposure in *Tsukamurella tyrosinosolvens* PS2

Romanova V.*, Kupriyanova E., Grigoryeva T., Laikov A.
Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia
**avonamora-94@mail.ru*

Key words: alkane; *alkB*; *P450*; alkane-degradation

Motivation and Aim: Microbial bioremediation is considered one of the most effective and cost-efficient approaches to remove hydrocarbons from contaminated ecosystems. The transformation and mineralization of organic pollutants by these microorganisms occurs due to the ability to use toxic substances to meet their energy and carbon needs. Most alkane-oxidizing bacteria contain a hydroxylase system, which is encoded by the *alkB* gene cluster. In addition, it has been shown that cytochrome *P450* is involved in the oxidation of alkanes with a chain length of C5-C16. There is some published data that alkane-oxidizing bacteria contain both *alkB* and *P450* systems [1], but it is not clear how they interact at the regulatory level and which one will play a dominant role in alkane degradation. The *T. tyrosinosolvens* PS2 strain is capable of using alkanes with different chain lengths as the sole carbon and energy source [2]. The present study aimed to evaluate the expression of alkane hydroxylase genes, such as *alkB* and *P450*, induced by alkanes of different chain lengths to elucidate the high biodegradability of PS2 strain.
Methods and Algorithms: Total RNA from PS2 strain cells was isolated with Qiagen RNeasy Mini kit according to the manufacturer's instruction with some modifications, under different cultivation conditions (sucrose, dodecane, hexadecane, hexatriacontane, liquid paraffin), cDNA was obtained, and real-time PCR was performed with specific primers to evaluate gene expression.
Results: Real-time PCR analysis showed that alkanes with different chain lengths can affect the two studied genes *(alkB* and *P450*), which leads to differences in expression between these genes when PS2 cells are grown under different culture conditions. Interestingly, on a medium supplemented with dodecane, the expression of the *P450* gene increased by 2,851 times, and the *alkB* gene by 1,229 times. The increase in P450 gene expression was also predominant in the medium supplemented with liquid paraffin and hexadecane as the sole energy source, compared to the *alkB* gene. On the other hand, on a medium supplemented with hexatriacontane, the expression of the *alkB* gene was increased by 32 times, and the *P450* gene by 7 times. Thus, a switch from the *P450* gene to the *alkB* gene is observed depending on the length of the hydrocarbon chain.
Conclusion: The obtained results demonstrated a preferential activation of the *P450* gene on medium-chain alkanes, although most studies describe, on the contrary, an increase in the expression of the *alkB* gene. The regulatory mechanisms identified in this study will be used to create tools to control cellular processes, increasing metabolic rates and the efficiency of biocatalysis.

Funding: This research was funded by the subsidy allocated to the Kazan Federal University for the state research assignment (No. FZSM-2023-0013) and the Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030).

References

1. Rojo F. Degradation of alkanes by bacteria. *Environ Microbiol*. 2009;11(10):2477-2490. doi 10.1111/j.1462-2920.2009.01948.x
2. Romanova V., Markelova M., Boulygina E., Siniagina M., Müller R., Grigoryeva T., Laikov A. Significance of both alkB and P450 alkane-degrading systems in Tsukamurella tyrosinosolvens: proteomic evidence. *Appl Microbiol Biotechnol*. 2022;106(8):3153-3171. doi 10.1007/s00253-022-11906-1

doi 10.18699/bgrs2024-8.2-15

Isolation and characterization of endolithic strain *Pseudomonas chlororaphis* S15: a study on heavy metal resistance and siderophore production

Shirshikova T.V., Ivoilova T.M., Elistratova A.A., Khilyas I.V.*

Kazan (Volga Region) Federal University, Institute of Fundamental Medicine and Biology, Kazan, Russia

** irina.khilyas@gmail.com*

Key words: endolithic bacteria; siderophores; *Pseudomonas chlororaphis*; heavy metals

Motivation and Aim: Serpentinite is a metamorphic rock resulting in the formation of serpentine minerals [1]. These minerals contain high levels of heavy metals, which provides a distinctive ecological niche for endolithic microbial communities. The endolithic bacteria found in serpentinite demonstrate a high resistance to heavy metals, the capacity to produce siderophores, and represent a valuable resource for understanding extremophile biology and for developing innovative biotechnological applications.

Endolithic bacteria have evolved unique adaptations to survive under extreme conditions, such as desiccation and nutrient limitation. The utilization of these bacteria for the improvement of soil in arid regions represents a promising strategy for the improvement of soil fertility and crop productivity. *Pseudomonas chlororaphis* is well known for its potential application in bioremediation, agriculture, and industrial biotechnology [2]. This bacterium produces different secondary metabolites, including pigments, antibiotics, and siderophores.

Thus, this study outlines the isolation and comprehensive characterization of serpentinite-derived strain *Pseudomonas chlororaphis* S15.

Methods and Algorithms:

Isolation and Identification: The strain S15 was isolated from a colony on Luria Agar (LA) agar that had been plated with aqueous rinsate of crushed mineral. Genomic DNA from strain S15 was extracted from an overnight LB-grown culture using the phenol-chloroform method. The 16S rRNA gene (1,500 bp) was amplified with polymerase chain reaction, then then sequenced using instrument ABI 3730 DNA Analyzer (Life Technologies, USA) following Sanger's method. The bacterial sequences were analyzed using the Basic Local Alignment Search Tool (BLASTn).

Heavy Metal Resistance: P. chlororaphis S15 was exposed to different concentrations of heavy metals (Ni^{2+}, Co^{2+}, Cu^{2+}, Zn^{2+}, Fe^{3+}, $Cr_2O_7^{2-}$) to determine their minimum inhibitory concentrations (MICs).

Siderophore Production Screening: The Chrome Azurol S (CAS) agar assay was used for screening *P. chlororaphis* S15 capacity for siderophore production [3]. Additionally, a replacement of Fe^{+3} with heavy-metal ions (Al^{+3}, Cu^{+2}, Ga^{+3}) in the standard CAS agar method was tested [4].

Growth Dynamics and Siderophore Production: The influence of different concentrations of heavy metals on *P. chlororaphis* S15 growth and siderophore production was studied in M9 minimal medium at 30 °C, 250 rpm. The Arnow and Atkin assay were used to detect catecholate- and hydroxamate-types of siderophores [5].

Metabolite Analysis: The solid phase extraction (SPE) of metabolites produced by *P. chlororaphis* S15 in the presence of heavy metal and BIP (2.2-bipiridyl) as a control was used. HPLC was performed on an Acclaim® PolarAdvantage II (PA2) C18 reverse-phase column (5 μm, 250 × 4.6 mm) using an UltiMate 3000 UHPLC system (Thermo Fisher Scientific, United States). Gradient elution was achieved using two mobile phases containing water with 0.01 % (vol/vol) trifluoroacetic acid (TFA) and 100 % (vol/vol) acetonitrile with 0.01 % (vol/vol) TFA at a flow rate of 1 mL/min. The elution of the siderophores was monitored at 220, 260, 285 nm and with a fluorescence detector (excitation at 470 nm and emission at 530 nm wavelength).

Results: The endolithic strain S15 was isolated from serpentinite sampled from the Khalilovsky massif, Russia [6]. The 16S rRNA sequencing was performed for identification of the strain. The BLAST results demonstrated that S15 exhibited a high degree of similarity to *Pseudomonas chlororaphis*, with a similarity score of 99 %. The resistance of *P. chlororaphis* S15 to various heavy metals was quantified by determining their minimum inhibitory concentrations (MICs). The tolerance of the strain to heavy metals was tested on LA supplemented with salts of heavy metals, and the results are presented in the Table 1.

Table 1. Growth of endolithic *P. chlororaphis* S15 in the presence of different concentrations (mM) of heavy metals

Bacterial strain	Ni^{2+}	Co^{2+}	$Cr_2O_7^{2-}$	Cu^{2+}	Zn^{2+}	Fe^{3+}
P. chlororaphis S15	3	1	0.25	4	4	5

The results demonstrated that *P. chlororaphis* S15 is capable of producing siderophores on CAS agar plates. The medium color changed from blue to yellow, indicating the presence of siderophores. The replacement of Fe^{+3} with Al^{+3}, Cu^{+2}, Ga^{+3} ions revealed the formation of increased halo zones formed around the *P. chlororaphis* S15. The analysis of siderophore accumulation in the culture supernatant over 120 hours of growth revealed that the presence of heavy metals in concentrations of 100 μM $CoCl_2 \times 6H_2O$, $AlCl_3$, $GaBr_3$ or $CuSO_4$ and 400 μM $NiSO_4 \times 7H_2O$ and $ZnSO_4$ inhibited the production of catecholate- and hydroxamate-types of siderophores. Starting form 24 hours of *P. chlororaphis* S15 growth, the M9 medium supplemented with 10 μM $GaBr_3$ stimulated the production of hydroxamate-type (600 ± 25 μM) and catechol-type (40 ± 5 μM) siderophores compared to the Fe-limited medium. The HPLC analysis of extracted metabolites produced by *P. chlororaphis* S15 revealed the presence of a metabolite with a retention time of 13.0 min exclusively in the supernatant, in the absence of any heavy metals. The 10 μM $GaBr_3$ concentration significantly stimulated the production of metabolites with RT = 16.3 and 17.7 min (Fig. 1).

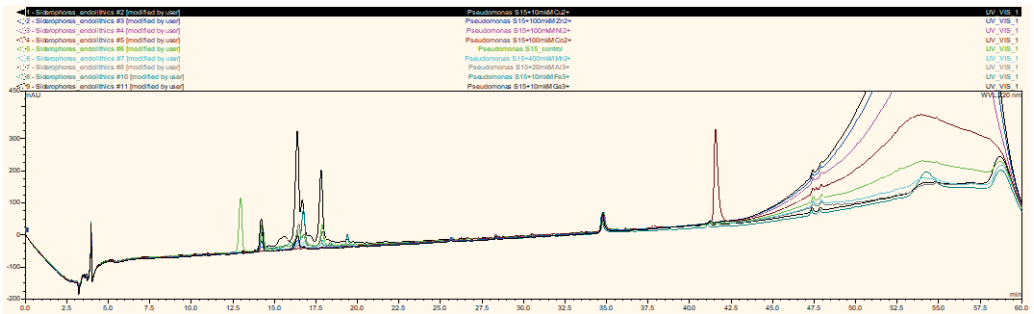


Fig. 1. HPLC analysis of extracted metabolites from the supernatant of *P. chlororaphis* S15 in the M9 medium without (green line) or supplemented with 10 μM $CuSO_4$ (brown line), 100 μM $ZnSO_4$ (blue line), 100 μM $NiSO_4 \times 7H_2O$ (pink line), 20 μM $CoCl_2 \times 6H_2O$ (red line), 400 μM $MnSO_4$ (turquoise line), 20 μM $AlCl_3$ (grey line), 10 μM FeCl3×6H2O (emerald line), 10 μM $GaBr_3$ (black line)

Conclusion: *Pseudomonas chlororaphis* is a promising bacterium with diverse applications in bioremediation, agriculture, and biotechnology. The capacity of *P. chlororaphis* S15 to synthesize siderophores and resist heavy metals renders it an optimal candidate for environmental remediation, enhancing plant health and suppressing pathogens. This underscores its potential for sustainable agriculture.

Funding: This research was supported by the RSF grant (Russian Science Foundation) No. 24-24-00473.

References

1. Sleep N.H. Geological and geochemical constraints on the origin and evolution of life. *Astrobiology.* 2018;18(9):1199-219. doi 10.1089/ast.2017.1778
2. Schwyn B., Neilands J.B. Universal chemical assay for the detection and determination of siderophores. *Analytical Biochemistry*. 1987;160(1):47-56. doi 10.1016/0003-2697(87)90612-9
3. Sarvepalli M., Velidandi A., Korrapati N. Optimization of siderophore production in three marine bacterial isolates along with their heavy-metal chelation and seed germination potential determination. *Microorganisms.* 2023;11(12):2873. doi 10.3390/microorganisms11122873
4. Payne S.M. Detection, isolation, and characterization of siderophores. *Methods in Enzymology*. 1994;235:329-344. doi 10.1016/0076-6879(94)35151-1
5. Raio A., Puopolo G. *Pseudomonas chlororaphis* metabolites as biocontrol promoters of plant health and improved crop yield. *World J Microbiol Biotechnol.* 2021;37(6):99. doi 10.1007/s11274-021-03063-w
6. Khilyas I.V., Sorokina A.V., Elistratova A.A., Markelova M.I., Siniagina M.N., Sharipova M.R., Shcherbakova T.A., D'Errico M.E., Cohen M.F. Microbial diversity and mineral composition of weathered serpentine rock of the Khalilovsky massif. *PLoS One.* 2019;14(12):e0225929. doi 10.1371/journal.pone.0225929

doi 10.18699/bgrs2024-8.2-16

The permafrost/active layer microbial community in the Eravna Bazin (Buryatia, Russia)

Zaitseva S.*, Kozyreva L., Badmaev N.
Institute of General and Experimental Biology, SB RAS, Ulan-Ude, Russia
** svet_zait@mail.ru*

Key words: microbial ecology; permafrost system; high-throughput Illumina sequencing

Motivation and Aim: Soil microbial communities play key role in biogeochemical cycles and greenhouse gas formation during the decomposition of the released organic matter in the thawing permafrost. The Eravna Basin, bounded by the Vitim Plateau in the northeast and the Selenga Middle Mountains in the south, is a model system for microbial ecology studies in various soil-ecological niches. The object of our studies was the Darkhituy–Khaimisan transect soils (the Eravna Basin, Buryatia). The soils of the transect polygon were well-characterized earlier [1, 2]. The active layer, which is a seasonally thawing surface and subsurface soil layer, suffers from repeated environmental disturbances due to frozen–thaw cycles [3, 4]. The active layer thickness differs across the sampling sites and was a significant factor determining bacterial and archaeal diversity. Active layer soils are exposed to multiple freeze thaw cycles during the change in seasons [4]. Environmental conditions of studied sites representing distinct habitats (larch forest, birch forest, meadow steppe and thermokarst lake) in the Eravna Basin characterized different amplitudes of freeze-thaw conditions over seasons and had a strong impact on taxonomical changes in the microbial communities present in the active layer during the initial period of soil thawing [5]. The aim of our research was to assess the taxonomic prokaryotic diversity at the seasonally thawing permafrost soils across Darkhituy-Khaimisan transect. We investigated changes in the microbial communities present in permafrost and the active layer of three sites representing distinct habitats. We explore the relationship between the biogeochemical differences among habitats and the permafrost/active layer microbial community via a spatial (across habitats, and with depth through the active layer) community survey using high-throughput Illumina sequencing.
Methods and Algorithms: We collected the seasonally thawed soil core and permafrost from the Darkhituy slope (1,015 m a.s.l.) and an adjacent thermokarst dried-up lake Khaimisan from the Eravna Basin (52° 38′ N, 111° 24′ E) during initial thawing phase in May, 2021, as well as the maximum thawing in September, 2022. Soil moisture, temperature, pH, soil organic carbon, soil nitrate and ammonia nitrogen were determined as previously described [5]. To isolate DNA from soil samples, was used a reagent kit (MACHEREY-NAGEL NucleoSpin Soil) from MACHEREY-NAGEL (Germany) according to the manufacturer's instructions. Library preparation and sequencing were carried out in accordance with the manufacturer's recommendations for operation on the Illumina MiSeq instrument (Illumina, USA) using the MiSeq® ReagentKit v3 (600 cycle) in a paired-end run (2*300 n). The research was done using equipment of the Core Centrum “GenomicTechnologies, Proteomics and Cell Biology” in ARRIAM (Saint-Petersburg, Russia). For the presentation of taxonomic analysis data, the tools of the QIIME software package [6] and SILVA software [7] were used. Statistical and

mathematical processing of multidimensional data of environmental parameters of the studied soils was carried out as previously described [5].

Results: The most common phyla across all samples of distinct permafrost thaw stages included *Proteobacteria* (34 %), *Actinobacteriota* (32 %), *Bacteroidota* (7 %), *Crenarchaeota* (7 %), *Firmicutes* (6 %), *Verrucomicrobiota* (4 %), *Acidobacteriota* (4 %), *Halobacterota* (1 %) and *Planctomycetota* (1 %). The relative abundance of dominant phyla *Crenarchaeota*, *Halobacterota*, *Verrucomicrobiota* and *Planctomycetota* as well as co-dominant taxa such as *Gemmatimonadota* and *Desulfobacterota* varied significantly ($p < 0.05$) between sample sites, with *Crenarchaeota*, *Verrucomicrobiota* and *Gemmatimonadota* abundant in the samples of the steppe meadows and the larch-birch forest and rare in lake Khaimisan soil samples, whereas Khaimisan soil samples were dominated by *Halobacterota*. Furthermore, the relative abundance of some of the dominant phyla varied significantly ($p < 0.05$) between Darkhitui soil horizons, with *Crenarchaeota*, *Acidobacteria*, *Planctomycetota*, *Verrucomicrobiota* and *Gemmatimonadota*, being more abundant in top-soil and subsurface soil horizon, whereas deep frozen horizons were dominated by *Actinobacteriota* and γ-*proteobacteria*. Specifically in Khaimisan permafrost samples, *Actinobacteriota*, with the families *Micrococcaceae*, *Microbacteriaceae* and *Nocardioidaceae* accounting for up to 37.4 % of reads in the permafrost samples and *Gammaproteobacteria* with family *Comamonadaceae* (11 %) and *Hydrogenophilaceae* (6.9 %) increased with depth. *Methanoperedenaceae*, *Devosiaceae*, *Hyphomicrobiaceae*, *Pyrinomonadaceae* and *Xanthomonadaceae* dominated during initial thawing in spring. *Beijerinckiaceae, Geodermatophilaceae*, *Intrasporangiacea* and *Micromonosporaceae* were abundant during maximal thawing in autumn. Significant differences ($p < 0.05$) of mean relative abundance at family level between initial and maximal thawing stages were found in Darkhituy meadow steppe. Thus, bacteria *Chthoniobacteraceae*, *Gemmatimonadaceae*, *Microbacteriaceae*, *Propionibacteriaceae, Pyrinomonadaceae,* unclassified_ *Rhizobiales, Xanthobacteraceae* and archaea *Nitrososphaeraceae* prevailed during initial thawing. Principal Coordinate Analysis (PCoA) of the beta diversity of soil microbial community revealed that microbial community structure on family level remained similar in the active layer depth regardless of the thawing stage. Top-soil horizons in May and in September were characterized by similar microbial community. While, microbial community structure in permafrost under birch forest varied significantly with domination of *Nocardioidaceae*, *Rhizobiaceae*, *Streptomycetaceae*. These three families accounted for more than 40 % of total microbial diversity. Similarly, analogous composition of microbial community was identified in the active layer under meadow steppe at different thawing stages. Microbial community structure varied significantly in permafrost layer at this sample site with high relative abundance of *Flavobacteriaceae*, *Oxalobacteraceae*, *Pseudomonadaceae*, *Sphingobacteriaceae*, *Comamonadaceae*, *Caulobacteraceae*, and *Methylophilaceae*. In the permafrost of the Khaimisan soil core, the abundance of gammaproteobacteria C*omamonadaceae* increased sharply (up to 23.9 %). The most abundant ASV demonstrates the greatest homology with strains of the psychrophilic, aerobic, chemoorganotrophic genera *Polaromonas* and *Variovorax*. Abundance of *Actinobacterota* increased toward permafrost depths under birch forest, reaching almost 51 % of the total abundance, mainly governed by the families *Nocardioidaceae*, *Streptomycetaceae*, *Micrococcaceae* and *Nocardiaceae*. Members of the genus *Streptomyces* make up almost 12 % of the permafrost community. The

increased abundance of *Bacteroidota* was found at permafrost under steppe. The taxa that dominated this 250-cm layer in our study included both the copiotrophic *Bacteroidia* and *Gammaproteobacteria*. Although *Bacteroidia* were found as one of the clearest thaw zone indicator taxa, we found its dominance (up to 36.6 %) in permafrost. *Flavobacterium*, *Pseudomonas*, *Massilia*, unclassified_ *Comamonadaceae* and *Pedobacter* were the five most abundant genera, which together account for 54 % of all sequences in permafrost.

Conclusion: A comprehensive analysis of the microbial distribution at the seasonally thawing permafrost soils of the Vitim Plateau revealed a site- and horizon-specific patterns of the microbial distribution at the seasonally thawing permafrost soils. High level of microbial diversity in all soil horizons of active layer during the maximum thawing and in intact permafrost was found. *Actinobacteriota* have been identified as one of the most obvious indicator taxa of permafrost horizons. *Proteobacteria* increased toward permafrost depths is mainly due to members of the copiotrophic family *Comamonadaceae*, which grow rapidly when supplied with fresh, labile nutrients.

Funding: The study is supported within the framework of the State Assignment No. 121030100229-1 for Institute of General and Experimental Biology, Siberian Branch, Russian Academy of Sciences.

References

1. Gyninova A. et al. Soils of the Darkhitui catena in the southern Vitim Plateau and their micromorphological features. *IOP Conf Ser Earth Environ Sci*. 2021;862:012068
2. Gyninova A. et al.On frozen chernozem genesis and evolution at lacustrine plains in the southern Vitim plateau (Eastern Siberia, Russia). *IOP Conf Ser Earth Environ Sci*. 2021;908:012033
3. Steven B. et al. Microbial diversity and activity through a permafrost/ground ice core profile from the Canadian high Arctic. *Environ Microbiol*. 2008;10:3388-3403
4. Chen Y.L. et al. Distinct microbial communities in the active and permafrost layers on the Tibetan Plateau. *Mol Ecol*. 2017;26:6608-6620
5. Zaitseva S. et al. Microbial community in permafrost thaw gradient in the south of the Vitim Plateau (Buryatia, Russia). *Microorganisms.* 2022;10(11):2202
6. Caporaso J.G. et al. QIIME allows analysis of highthroughput community sequencing data. *Nat Methods.* 2010;7:335-336
7. Quast C. et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(Database issue):D590-D596

8 Симпозиум «Микробиология и биотехнологии: компьютерные и экспериментальные подходы»

Symposium “Microbiology and biotechnologies: computational and experimental approaches”

doi 10.18699/bgrs2024-8.3-01

Поиск генов катаболизма стероидов и автоматизация работы программы BLAST+

Брагин Е.Ю.*, Фокина В.В., Донова М.В.

Институт биохимии и физиологии микроорганизмов им. Г.К. Скрябина Российской академии наук, обособленное подразделение Федерального исследовательского центра «Пущинский научный центр биологических исследований Российской академии наук», Пущино, Россия

** bragory@yandex.ru*

Ключевые слова: стероиды; Перл; программа; поиск генов; бактерии

Мотивация и цель: Развитие технологий полногеномного секвенирования открыло перспективу осуществления *in silico* поиска организмов, обладающих определенными свойствами. Такой поиск может осуществляться с помощью программы BLAST+. Работа с данной программой может быть относительно легко автоматизирована с помощью мини-программ, написанных на языке программирования Perl. Одной из областей, в которых подобный автоматизированный BLAST+ поиск может быть применим, является поиск бактерий, обладающих ферментами катаболизма стероидов. Стероидные соединения активно используются в различных областях медицины, а также ветеринарии. При получении подобных соединений активно используются биотехнологические методы. Штаммы микроорганизмов, применяемые в биотехнологических процессах, в свою очередь, могут быть получены с помощью методов генной инженерии из бактерий, обладающих ферментами катаболизма стероидов.

Методы и алгоритмы: Для автоматизации работы программы BLAST+ на языке Perl была написана программа blastconductor.pl. Данная программа использует нуклеотидные последовательности геномов и аминокислотные последовательности референсных белков. Она запускает выравнивание каждой геномной последовательности с каждой референсной последовательностью с помощью программы BLAST+ в режиме tblastn. Результат каждого такого выравнивания сохраняется в виде временного файла. Из каждого такого файла программа извлекает информацию о длине референсной последовательности; количество идентичных аминокислот для каждого выровнявшегося участка; e-value, и др. Далее происходит отбор и анализ выровнявшихся последовательностей. Результаты сохраняются в виде отдельных файлов. Для поиска генов катаболизма стероидов, помимо описанной выше программы blastconductor.pl, использовали программу BLAST+ версии 2.11.0, а также аминокислотные последовательности 230 референсных белков, которые связаны с катаболизмом стеринов (111 последовательностей), желчных кислот (46 последовательностей), а также диосгенина (73 последовательности). Поиск генов, ассоциированных с катаболизмом стероидов, проводили среди 567 известных геномов прокариот, депонированных во Всероссийской коллекции микроорганизмов (ВКМ) в ИБФМ РАН. Все референсные нуклеотидные и аминокислотные последовательности были получены из базы данных GenBank (NCBI).

Результаты: Поиск генов, связанных с катаболизмом стероидов, позволил выявить в 64 бактериальных геномах не менее 30 генов, схожих с генами, ассоциированными с катаболизмом стеринов не менее чем на 50 %. Из них 21 относились к роду *Rhodococcus*, 16 – к роду *Streptomyces,* 6 штаммов – к роду *Gordonia*, 4 – к роду *Actinomadura* и 3 штамма – к роду *Amycolatopsis.* В этой группе также присутствовали представители родов *Mycolicibacterium*, *Nocardia*, *Saccharopolyspora*, *Couchioplanes*, *Nocardioides*, *Actinocorallia*, *Actinoplanes* и *Crossiella*. От 20 до 29 таких генов было обнаружено у большого количества штаммов (табл. 1). Количество штаммов, у которых найдено значительное количество генов, схожих с генами катаболизма желчных кислот, было значительно меньше (см. табл. 1). Меньше всего было обнаружено штаммов, обладающих значительным количеством генов, схожих с генами кластера катаболизма диосгенина (см. табл. 1).

Таблица 1. Количество выявленных штаммов с различным количеством генов, ассоциированных с катаболизмом стероидных соединений

Количество генов, схожих с референсными генами не менее чем на 50 %	Количество бактериальных штаммов, у которых выявлены гены, схожие с генами катаболизма		
	стеринов	желчных кислот	диосгенина
10–19	20	3	23
20–29	98	18	3
>30	64	15	1

Выводы: Создана автоматизирующая работу BLAST+ программа blastconductor.pl, позволяющая быстро и просто осуществлять *in silico* поиск значительного количества генов в большом количестве геномов. Найдено значительное количество бактериальных штаммов, обладающих генами, связанными с катаболизмом стероидов, которые могут быть использованы при создании штаммов, задействованных в получении стероидных соединений.

Финансирование: Работа выполнена при финансовой поддержке Российского научного фонда (проект № 21-64-00024).

The search for steroid catabolism genes and automation of the BLAST+ program

Bragin E.Yu.*, Fokina V.V., Donova M.V.

G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences; Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, Russia

** bragory@yandex.ru*

Key words: steroids; Perl; program; gene search; bacteria

Motivation and Aim: The full-genome sequencing technologies development had opened up the possibility of *in silico* searching for organisms with certain properties. Such a search can be carried out using the BLAST+ program. Working with this program can

be relatively easily automated using mini-programs written in the Perl programming language. One of the areas in which such an automated BLAST+ search can be applied is the search for bacteria possessing steroid catabolism enzymes. Steroid compounds are actively used in various fields of medicine, as well as in veterinary. Biotechnological methods are actively used in the production of such compounds. Strains used in biotechnological processes can be obtained using genetic engineering methods from bacteria possessing steroid catabolism enzymes.

Methods and Algorithms: To automate the operation of the BLAST+ program, the program which was named blastconductor.pl was written in Perl language. This program uses nucleotide sequences of genomes and amino acid sequences of reference proteins. It starts the alignment of each genomic sequence with each reference sequence using the BLAST+ program in tblastn mode. The result of each alignment is saved as a temporary file. From each such file the program extracts the following information: length of the reference sequence; number of identical amino acids for each aligned region; *e*-value and others. The selection and analysis of aligned sequences are carried out. The results are saved as separate files. To search for steroid catabolism genes, in addition to the blastconductor.pl program described above, the BLAST+ program version 2.11.0 was used, as well as the amino acid sequences of 230 reference proteins, which associated with catabolism of sterols (111 sequences), bile acids (46 sequences) and diosgenin (73 sequences). An *in silico* search for genes associated with steroid catabolism was carried out among 567 known prokaryotic genomes deposited in All-Russian collection of microorganisms (VKM) IBFM RAS. All reference nucleotide and amino acid sequences were obtained from the GenBank database (NCBI).

Results: The search for genes associated with steroid catabolism made it possible to identify at least 30 genes in 64 bacterial genomes that are similar to known genes associated with sterol catabolism by at least 50 %. Of these, 21 belong to the genus *Rhodococcus*, 16 to the genus *Streptomyces*, 6 strains to the genus *Gordonia*, 4 to the genus *Actinomadura* and 3 strains to the genus *Amycolatopsis*. Representatives of the genera *Mycobacterium*, *Nocardia*, *Saccharopolyspora*, *Couchioplanes*, *Nocardioides*, *Actinocorallia*, *Actinoplanes* and *Crossiella* were also present in this group. From 20 to 29 such genes were found in a large number of strains (Table 1). The number of strains in which a significant number of genes similar to the genes of the bile acid catabolism were found was significantly smaller (Table 1). The least number of strains were found with a significant number of genes similar to the genes associated with the diosgenin catabolism (Table 1).

Table 1. The amount of founded strains with different numbers of genes associated with the catabolism of sterols, bile acids and diosgenin

The number of genes similar to the reference genes by at least 50 %	The number of revealed bacterial strains possessing genes similar to genes associated with the catabolism of		
	sterol	bile acids	diosgenin
10–19	20	3	23
20–29	98	18	3
> 30	64	15	1

Conclusion: The program, which automates the operation of BLAST+, has been created to quickly and easily perform *in silico* searches for a significant number of genes, in a large amount of genomes. A significant number of bacterial strains have been found that possessing genes associated with steroid catabolism, which can be used in the creation of strain-biocatalysts involved in the production of valuable steroid compounds.

Funding: This work was financially supported by the Russian Science Foundation (project No. 21-64-00024).

doi 10.18699/bgrs2024-8.3-02

Коллекция микроорганизмов ИЦиГ СО РАН как основа для создания биотехнологически перспективных штаммов-продуцентов

Брянская А.В.[1, 2]*, Уварова Ю.Е.[1, 2], Горячковская Т.Н.[1, 2], Шипова А.А.[1, 2], Задорожный А.В.[1, 2], Богачева Н.В.[1, 2], Коржук А.В.[1, 2], Шляхтун В.Н.[1, 2], Мещерякова И.А.[1, 2], Чесноков Д.О.[1, 2], Павлова Е.Ю.[1, 2], Бочков Д.В.[1, 2], Банникова С.В.[1, 2], Васильева А.Р.[1, 2], Пельтек С.Е.[1, 2]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

* *alla@bionet.nsc.ru*

Ключевые слова: коллекция микроорганизмов; штамм-продуцент; биотехнология

Мотивация и цель: Создание эффективных штаммов-продуцентов невозможно без прочного фундамента в виде коллекции микроорганизмов, обладающих биотехнологическими свойствами. Для этой цели помимо направленного поиска, сбора и выделения микроорганизмов, необходимо всестороннее их изучение с использованием классических и современных инструментов, методов и подходов, в частности протеомики, транскриптомики, метаболомики и биоинформатики, которая позволяет провести анализ полученных данных для поиска и создания новых эффективных штаммов-продуцентов.

Методы и алгоритмы: Коллекция микроорганизмов создавалась в результате отбора образцов в экстремальных по температуре, pH, солености, уровню загрязнения и другим параметрам экосистемах, расположенных на территории Российской Федерации. Выделение микроорганизмов проводилось классическими микробиологическими методами. Для штаммов коллекции проведено полногеномное секвенирование на приборе Illumina MiSeq (США) с последующим аннотированием. Поиск соответствующих генов и ферментов, определение метаболомных и протеомных профилей штаммов проведен с использованием комплекса методов протеомики, метаболомики и биоинформатики.

Результаты: В рамках данной работы создана и постоянно пополняется новыми культурами Коллекция микроорганизмов биотехнологического назначения ФИЦ ИЦиГ СО РАН. В настоящее время в коллекции имеется более 3000 культур, среди которых есть не только представители различных филумов, но и доменов. В геномах штаммов найдены гены целевых ферментов: амилазы, протеазы, маннаназы, ксиланазы, целлюлазы, необходимые для осуществления современных биотехнологических процессов. На этой основе проведено клонирование целевых генов в геном дрожжей *Komagataella phaffii* T07 и получены эффективные штаммы-продуценты и препараты ферментов амилазы, маннаназы, ксиланазы, протеазы, целлюлазы и др.

Выводы: Для создания эффективных штаммов-продуцентов нужна последовательность определенных действий: поиск, выделение и характеризация исходных

штаммов, а затем непосредственно генно-инженерные и биотехнологические работы. В Коллекции микроорганизмов биотехнологического назначения ФИЦ ИЦиГ СО РАН осуществляется именно такой комплексный подход, который позволил сформировать прочную базу для генетических, молекулярно-биологических и биотехнологических исследований и позволяет в настоящее время получать эффективные штаммы-продуценты и препараты ферментов.
Финансирование: Исследование поддержано Министерством науки и высшего образования Российской Федерации (№ FWNR-2022-0022) и «Курчатовским геномным центром ИЦиГ СО РАН» (№ 075-15-2019-1662).

Collection of microorganisms of FRC ICG SB RAS as a basis for creation of biotechnologically promising strains-producers

Bryanskaya A.V.[1, 2]*, Uvarova Y.E.[1, 2], Goryachkovskaya T.N.[1, 2], Shipova A.A.[1, 2], Zadorozhny A.V.[1, 2], Bogacheva N.V.[1, 2], Korzhuk A.V.[1, 2], Shlyakhtun V.N.[1, 2], Mescheryakova I.A.[1, 2], Chesnokov D.O.[1, 2], Pavlova E.Yu.[1, 2], Bochkov D.V.[1, 2], Bannikova S.V.[1, 2], Vasilieva A.R.[1, 2], Peltek S.E.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *alla@bionet.nsc.ru*

Key words: microbial collection; strains-producers; biotechnology

Motivation and Aim: Creation of effective strains-producers is impossible without a solid foundation in the form of a collection of microorganisms with biotechnological properties. For this purpose, in addition to purposeful search, collection and isolation of microorganisms, it is necessary to study them comprehensively using classical and modern tools, methods and approaches, in particular proteomics, transcriptomics, metabolomics and bioinformatics, allowing to analyse the obtained data in order to search for and create new effective strains-producers.
Methods and Algorithms: The collection of microorganisms was created as a result of sampling in extreme ecosystems of the Russian Federation in terms of temperature, pH, salinity, pollution level and other parameters. Microorganisms were isolated using classical microbiological methods. Full genome sequencing on Illumina MiSeq (USA) with subsequent annotation was performed for the strains of the collection. The search for relevant genes and enzymes, determination of metabolomic and proteomic profiles of strains was carried out using a combination of proteomics, metabolomics and bioinformatics methods.
Results: The collection of microorganisms for biotechnological purposes of FRC ICG SB RAS has been established and is constantly replenished with new cultures. Currently, the collection includes more than 3,000 cultures, among which there are representatives of not only different phyla, but also domains. Genes of target enzymes: amylase, protease, mannanase, xylanase, cellulase, necessary for modern biotechnological processes, were found in the genomes of the strains. On this basis, the target genes were cloned into the genome of yeast *Komagataella phaffii* T07 and effective strains-producers and preparations of enzymes amylase, mannanase, xylanase, protease, cellulase, etc. were obtained.

Conclusion: Creation of effective producer strains requires a sequence of certain actions: search, isolation and characterisation of initial strains, and then direct genetic engineering and biotechnological work. The collection of microorganisms for biotechnological purposes of FRC ICG SB RAS implements just such an integrated approach, which has allowed to form a solid base for genetic, molecular-biological and biotechnological research and allows to obtain effective strains-producers and enzyme preparations.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (No. FWNR-2022-0022) and Kurchatov Genomic Center, Federal Research Center ICG SB RAS (No. 075-15-2019-1662).

doi 10.18699/bgrs2024-8.3-03

In silico поиск и анализ мутантных форм повышенной термостабильности промышленно значимого фермента савиназы из *Bacillus lentus* путем оптимизации его поверхностных заряд-зарядовых взаимодействий

Величко А.А.*, Миловидов Г.Д., Дерюшева Е.И., Шевелёва М.П., Немашкалов В.А.

ФГБУН ФИЦ «Пущинский научный центр биологических исследований» РАН, Институт биологического приборостроения, Пущино, Московская область, Россия

** vaconf13@yandex.ru*

Ключевые слова: савиназа; электростатические взаимодействия; термостабильность; TKSA-MC; AlphaFold

Мотивация и цель: Ферменты широко применяются в различных промышленных отраслях для обработки первичного сырья. По сравнению с традиционными химическими процессами биохимические подходы к осуществлению технологических процессов позволяют добиться большего выхода целевого продукта за счет высокой каталитической активности и специфичности ферментов, а также экономических и экологических преимуществ за счет более мягких условий, уменьшения объемов и токсичности отходов. Протеазы, вызывающие расщепление пептидной связи и гидролизующие белки до небольших пептидов и аминокислот, используются для обработки кожи и тканей, улучшения усвояемости кормов и качества пищи, при производстве моющих средств и биотоплива. Повышение их стабильности и активности в промышленных условиях их функционирования является одной из главных задач модификации этого класса белков. Фермент савиназа, секретируемая алкалафильной бактерией *Bacillus lentus,* относится к сериновым протеазам и в настоящее время широко используется в моющих средствах в качестве добавки для удаления белковых отложений. Её оптимальный рабочий диапазон температур 30–70 °C и pH 8–12. Белок связывает два иона кальция, которые наряду с семью солевыми мостиками между ионизированными аминокислотными остатками способствуют стабилизации его структуры [1]. Стабильность савиназы можно модулировать путем добавления в раствор осмолитов [2], но этот подход не является универсальным, дает ограниченный эффект и не всегда подходит для промышленного использования.

Для повышения стабильности белков методами белкового инжиниринга обычно используют оптимизацию внутримолекулярных взаимодействий, например, через переупаковку гидрофобного ядра или создание новых дисульфидных, водородных или ионных связей. Кроме того, повышение стабильности белов можно осуществить посредством оптимизации поверхностных заряд-зарядовых взаимодействий. Данный подход был успешно апробирован, например, для фермента ксиланаза из *Bacillus subtilis* [3].

В данной работе был проведен *in silico* поиск и анализ мутантных форм повышенной термостабильности савиназы путём оптимизации ее поверхностных заряд-зарядовых взаимодействий.

Методы и алгоритмы: Структура савиназы (UniProt ID: P29600) была взята из Protein Data Bank (https://www.rcsb.org/), PDB ID: 1SVN. Вклад в изменение энергии Гиббса в процессе денатурации белка, обусловленный заряд-зарядовыми взаимодействиями, ΔG_{qq}, вычислялся с помощью программы TKSA-MC (*T*anford-*K*irkwood *S*olvent *A*ccessibility model with the *M*onte *C*arlo method) (https://github.com/contessoto/tksamc) при pH 8.0 и 27 °C [4] как полная энергия электростатического взаимодействия отдельных пар ионизированных аминокислотных остатков на основании статистической суммы всех возможных ионизированных состояний при данном pH. При этом предполагалось, что в полностью развернутом состоянии электростатическое взаимодействие отсутствует. На дестабилизирующий вклад в структуру белка ионизированных аминокислотных остатков указывает значение вклада данной пары остатков *i* и *j* $\Delta E_{ij} > 0$, а также $SASA_{ij} > 50$ % (*S*olvent-*A*ccessible *S*urface *A*rea – средняя площадь поверхности пары остатков, доступная растворителю) [5]. Структуры мутантных форм белка были получены с помощью программы AlphaFold 2 (https://www.alphafold.ebi.ac.uk) [6].

Результаты: Анализ профиля ΔG_{qq}, полученного для савиназы с помощью TKSA-MC, показал, что аминокислотные остатки Lys229, Lys231 и Lys245 на поверхности белковой молекулы вносят вклад в ее дестабилизацию. Кроме того, электростатическое отталкивание между остатками лизина с одной стороны и остатками гистидина и аргинина с другой (Lys229:His118; Lys231:Arg269; Lys245:Arg241; Lys245:His243) понижает стабильность белковой молекулы. Инверсия заряда путём замены этих остатков на отрицательно заряженную аспарагиновую кислоту (Asp) должна способствовать повышению стабильности фермента, что количественно выражается в изменении ΔG_{qq} в результате мутации $\Delta\Delta G_{wt\rightarrow mut} = \Delta G_{mut} - \Delta G_{wt}$. Анализ мутации Lys229Asp показал, что она не вносит существенного вклада в изменение стабильности белка ($\Delta\Delta G_{Lys229Asp} = 0.77$ кДж/моль). Замена Lys231Asp приводит к дестабилизации савиназы ($\Delta\Delta G_{Lys231Asp} = 14.49$ кДж/моль), в то время как мутация Lys245Asp приводит к повышению стабильности фермента ($\Delta\Delta G_{Lys245Asp} = -1.70$ кДж/моль). Двойные мутации не способствуют повышению стабильности савиназы (Lys229Asp+Lys231Asp: $\Delta\Delta G_{229+231} = 2.52$ кДж/моль; Lys231Asp+Lys245Asp: $\Delta\Delta G_{231+245} = 4.92$ кДж/моль). Добавление точечной замены Lys229Asp к Lys245Asp не влияло на значение $\Delta\Delta G$ по сравнению с одиночной заменой в положении 245 ($\Delta\Delta G_{229+245} = -1.70$ кДж/моль). Замена всех трёх остатков (Lys229 Asp, Lys231Asp, Lys245Asp) способствовала значительному повышению стабильности белковой молекулы: $\Delta\Delta G_{229+231+245} = -7.98$ кДж/моль. Отметим, что проведенные точечные мутации также привели к более «выгодным» электростатическим взаимодействиям с соседними положительно заряженными аминокислотами. Для аминокислотных остатков His118 и His243, расположенных в трехмерной структуре близко к Asp229 и Asp245, значение ΔG изменилось с 0.00 до –4.88 кДж/моль и до –3.97 кДж/моль соответственно, что также устранило дестабилизирующие взаимодействия.

Выводы: Таким образом, в данной работе был проведен анализ вклада поверхностных электростатических взаимодействий заряженных

аминокислотных остатков савиназы. Были выбраны точечные замены, потенциально приводящие к повышению стабильности фермента. Полученные данные будут в дальнейшем использованы для дизайна и анализа мутантных форм савиназы повышенной термостабильности.

In silico search and analysis of mutant forms of increased thermostability of the industrially important enzyme savinase from *Bacillus lentus* by optimizing its surface charge-charge interactions

Velichko A.A.*, Milovidov G.D., Deryusheva E.I., Shevelyova M.P., Nemashkalov V. A.

Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences, Institute for Biological Instrumentation of the Russian Academy of Sciences, Pushchino, Russia

** vaconf13@yandex.ru*

Key words: savinase; electrostatic interactions; thermal stability; TKSA-MC; AlphaFold

Motivation and Aim: Enzymes are widely used in various industrial sectors for processing raw materials. Compared to traditional chemical processes, biochemical approaches to the implementation of technological processes make it possible to achieve a higher yield of the target product due to high catalytic activity and specificity of enzymes, as well as economic and environmental advantages due to milder conditions and reduction in toxicity and volume of waste. Proteases, which cause cleavage of peptide bonds and hydrolyze proteins into small peptides and amino acids, are used in treatment of leather and fabrics, improving of feed digestibility and food quality, and in the production of detergents and biofuels. Increasing their stability and activity under industrial conditions is one of the main tasks of modifying this class of proteins. The enzyme savinase is a serine protease secreted by the alkaliphilic bacterium *Bacillus lentus* and is widely used in detergents as an additive to remove protein stains. Its optimal operating temperature range is 30–70 °C and pH 8–12. The protein binds two calcium ions, which, along with seven salt bridges between ionized amino acid residues, help stabilize its structure [1]. The stability of savinase can be modulated by adding osmolytes to the solution [2], but this approach is not universal, has a limited effect and is not always suitable for industrial use.

To increase the stability of proteins, engineering methods usually use optimization of intramolecular interactions, for example, through repacking of the hydrophobic core or the creation of new disulfide, hydrogen or ionic bonds. The increase of protein stability can also be achieved by optimizing surface charge-charge interactions. This approach has been successfully tested, for example, for the xylanase from *Bacillus subtilis* [3].

In this work, an *in silico* search and analysis of mutant forms of increased thermal stability of savinase was carried out by optimizing its surface charge-charge interactions.

Methods and Algorithms: The structure of savinase (UniProt ID: P29600) was taken from the Protein Data Bank (https://www.rcsb.org/), PDB ID: 1SVN. The contribution to the Gibbs energy change during protein denaturation caused by charge-charge interactions, ΔG_{qq}, was calculated using the TKSA-MC program (*T*anford-*K*irkwood

*S*olvent *A*ccessibility model with the *M*onte *C*arlo method) (https://github.com/contessoto/tksamc) at pH 8.0 and 27 °C [4] as the total energy of electrostatic interaction of individual pairs of ionized amino acid residues based on the partition function of all possible ionized states at a given pH. It was assumed that in the fully denatured state there is no electrostatic interaction. The destabilizing contribution of ionized amino acid residues to the protein structure is indicated by the contribution of a given pair of residues *i* and *j* $\Delta E_{ij} > 0$, as well as $SASA_{ij} > 50$ % (*S*olvent-*A*ccessible *S*urface *A*rea) [5]. The structures of mutant forms of the protein were obtained by AlphaFold 2 program (https://www.alphafold.ebi.ac.uk) [6].

Results: Analysis of the ΔG_{qq} profile obtained for savinase using TKSA-MS showed that the amino acid residues Lys229, Lys231 and Lys245 on the protein surface contribute to its destabilization. In addition, electrostatic repulsion between pairs of lysine residues with histidine or arginine residues (Lys229:His118; Lys231:Arg269; Lys245:Arg241; Lys245:His243) reduces the stability of the protein molecule. Charge inversion by replacing these residues with negatively charged aspartic acid (Asp) should increase the stability of the enzyme, which is quantified by the change in ΔG_{qq} because of the mutation $\Delta\Delta G_{wt\rightarrow mut} = \Delta G_{mut} - \Delta G_{wt}$. Analysis of the Lys229Asp mutation showed that it does not significantly contribute to protein stability ($\Delta\Delta G_{Lys229Asp} = 0.77$ kJ/mol). The Lys231Asp substitution leads to destabilization of savinase ($\Delta\Delta G_{Lys231Asp} = 14.49$ kJ/mol), while the Lys245Asp mutation leads to increased enzyme stability ($\Delta\Delta G_{Lys245Asp} = -1.70$ kJ/mol). Double mutations do not increase the stability of savinase (Lys229Asp+Lys231Asp: $\Delta\Delta G_{229+231} = 2.52$ kJ/mol; Lys231Asp+Lys245Asp: $\Delta\Delta G_{231+245} = 4.92$ kJ/mol). The addition of a substitution Lys229Asp to Lys245Asp had no effect on the $\Delta\Delta G$ value compared to a point substitution at position 245 ($\Delta\Delta G_{229+245} = -1.70$ kJ/mol). The replacement of all the residues (Lys229 Asp, Lys231Asp, Lys245Asp) led to a significant increase in the stability of the protein molecule: $\Delta\Delta G_{229+231+245} = -7.98$ kJ/mol. Note that the point mutations also led to more "favorable" electrostatic interactions with neighboring positively charged amino acids. For amino acid residues His118 and His243, located in the tertiary structure close to Asp229 and Asp245, the ΔG value changed from 0.00 to –4.88 kJ/mol and to –3.97 kJ/mol, respectively, which also eliminated destabilizing interactions.

Conclusion: Thus, in this work, the contribution of surface electrostatic interactions of charged amino acid residues of savinase was analyzed. Point substitutions that potentially lead to increased enzyme stability were selected. The data obtained will be further used for the design and analysis of mutant forms of savinase with increased thermostability.

Список литературы/References

1. Betzel C. et al. Crystal structure of the alkaline proteinase Savinase™ from Bacillus lentus at 1.4 Å resolution. *J Mol Biol*. 1992;223(2):427-445
2. Nasiripourdori A. et al. Co-solvent effects on structure and function properties of savinase: Solvent-induced thermal stabilization. *Int J Biol Macromol*. 2009;44(4):311-315
3. Ngo K. et al. Improving the Thermostability of Xylanase A from *Bacillus subtilis* by Combining Bioinformatics and Electrostatic Interactions Optimization. *J Phys Chem B*. 2021;125(17):4359-4367
4. Contessoto V.G. et al. TKSA-MC: A web server for rational mutation through the optimization of protein charge interactions. *Proteins*. 2018;86(11):1184-1188
5. Loladze V.V. et al. Engineering a Thermostable Protein via Optimization of Charge − Charge Interactions on the Protein Surface. *Biochemistry*. 1999;38(50):16419-16423
6. Mirdita M. et al. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022;19(6):679-682

doi 10.18699/bgrs2024-8.3-04

Ортогональная трансляция, изменение специфичности аминоацил тРНК синтетаз

Розанов А.*, Дахневич А., Илюшин Д., Мясоутова С., Тихонова Е., Кукушкина А., Иванов Р.

*Научно-технологический университет «Сириус», **федеральная территория «Сириус»**, Россия*
** rozanov.as@talantiuspeh.ru*

Ключевые слова: аминоацил тРНК синтетаза; неприродная/непротеиногенная аминокислота; направленная эволюция; пирролизин

Мотивация и цель: Один из путей создания белков и пептидов с новыми свойствами – это включение аминокислот, не встречающихся в природе. Такие аминокислоты называются неприродными или непротеиногенными (НПА). Традиционно такие аминокислоты встраивают, обычно в небольшие пептиды, химическим синтезом или посттрансляционной модификацией аминокислотного остатка. Этот подход имеет множество ограничений, в том числе экономические. Включение НПА позволяет вводить химические группы, которые обеспечивают недоступные для природных белков свойства. В том числе возможна прижизненная встройка НПА [1].

Для чего это может быть нужно? При помощи НПА со взаимно комплементарными группами возможно специфически соединять белки и пептиды, направленно присоединять органические соединения, в обычных условиях без защиты других аминокислотных остатков. Возможно вносить защищенные или модифицированные групировки для химических преобразований. Прежде всего, такой подход может быть актуален при разработке лекарственных средств. Для изучения влияния посттрансляционных модификаций возможно также включение аминокислот уже с модификациями. Отдельные НПА могут иметь флюоресцирующие, или регистрируемые другими методами, группы, что может быть интересно для прижизненного исследования биологических объектов. Еще один вариант использования – это модификация функциональных атомов для изменения аффинности или реакционной способности. Цель научной группы в НТУ Сириус – в первую очередь включение НПА для конструирования пептидов и белков, обладающих уникальными свойствами, для создания фармацевтических субстанций.

В настоящее время пептиды с НПА, встроенными при помощи биологического синтеза, не получили широкого признания, тогда как НПА, встроенные химическим синтезом, находят широкое применение. Это говорит о перспективности данного направления.

Методы и алгоритмы: Первым шагом в развитии биологических систем, направленных на получение аминокислотных последовательностей с НПА, является получение пар аминоацил тРНК синтетаза(АРС)/тРНК, ортогональных другим аминокислотам, тРНК и АРС клетки. АРС не взаимодействует с другими тРНК и аминокислотами. тРНК не взаимодействует с другими АРС в клетке. При этом тРНК должна взаимодействовать с элементами системы трансляции рибосомой и факторами элонгации.

Помимо 20 аминокислот, которые встраиваются в белки большинством организмов, существуют еще две – селеноцистеин и пирролизин. Селеноцистеин включается с использованием дополнительных механизмов, а пирролизин – традиционного механизма рибосомального синтеза при помощи дополнительной пары АРС/тРНК. Данная пара была обнаружена в геноме архей и оказалась ортогональной к хозяйским парам АРС/тРНК как бактерий, так и эукариот [2]. Именно эта пара стала донором для получения достаточно широкого разнообразия НПА в белки и пептиды. В настоящее время продемонстрировано включение более 200 вариантов НПА в белки после направленных изменений. В основном это производные лизина и фенилаланина/тирозина [3].

Несмотря на то что была показана возможность включения широкого спектра НПА, в большинстве случаев эффективность включения довольно низка. Требуется дополнительная работа для повышения как селективности включения конкретных НПА, так и эффективности работы АРС. Возможны два основных принципа для изменения свойств АРС. Первый принцип – это рациональный дизайн. Второй – направленная эволюция. Мы использовали второй вариант – метод PACE (phage assisted continues evolution) [4].

Результаты: При помощи PACE из пары пирролизиновой АРС/тРНК были получены варианты, способные включать ряд аминокислот производных тирозина: О-аллил-тирозин, о-триазид-тирозин, о-метил-тирозин, п-амин-фенилаланин и 2-хлор-фенилаланин. На первом этапе направленную эволюцию АРС проводили в присутствии 10 мМ целевых аминокислот. При меньшей концентрации первый этап эволюции не удавалось проводить. В ходе последовательных итераций направленной эволюции удалось добиться работы системы при снижении концентрации целевых НПА до 2.5 мМ. Работы по повышению эффективности АРС продолжаются.

Для анализа свойств полученных АРС они были клонированы в штаммы *E. coli*, экспрессирующей ген белка GFP с амбер стоп-кодоном в 6-м положении. Эффективность работы оценивали по уровню флюоресценции. Масс-спектрометрический анализ GFP с включенным о-аллил-тирозином показал, что варианты АРС из начальных этапов PACE обладают не только меньшей эффективностью, но и меньшей селективностью. То есть помимо целевой аминокислоты могут включать природные аминокислоты тирозин и фенилаланин. Варианты АРС с более поздних этапов эксперимента обладали значительно большей специфичностью.

Выводы: Впервые в России получены штаммы продуценты белков с НПА. Таким образом, нами отработан протокол получения ортогональных пар АРС/тРНК для включения целевых НПА производных фенилаланина/тирозина. Данный протокол может быть расширен на производные лизина, гистидина и триптофана. Ввиду того, что в качестве кодона мы можем использовать один вариант – амбер стоп-кодон, в настоящий момент возможно получение пептидов и белков с одним вариантом НПА. Расширение возможностей для параллельного включения нескольких НПА требует разработки нескольких взаимно ортогональных пар и работы над другими элементами транскрипции и трансляции.

Финансирование: Исследование поддержано в рамках реализации государственной программы федеральной территории «Сириус» «Научно-технологическое развитие федеральной территории «Сириус».

Orthogonal translation, changing the specificity of aminoacyl tRNA synthetases

Rozanov A.*, Dakhnevich A., Ilyushin D., Myasoutova S., Tikhonova E., Kukushkina A., Ivanov R.
Sirius University, Sirius, Russia
** rozanov.as@talantiuspeh.ru*

Key words: aminoacyl tRNA synthetase; non-natural/non-canonic amino acid; directed evolution; pyrrolysine

Motivation and Aim: One way to create proteins and peptides with new properties is to incorporate amino acids not found in nature. Such amino acids are called non-natural or non-canonic amino acids (ncAA). Traditionally, such amino acids are incorporated, usually into small peptides, by chemical synthesis or post-translational modification of an amino acid residue. This approach has many limitations, including economic ones. The incorporation of ncAAs allows the introduction of chemical groups that provide properties unavailable to natural proteins. It is also possible to incorporate ncAAs in life cells or animals [1].

What can it be used for? With the help of ncAAs with mutually complementary groups it is possible to specifically connect proteins and peptides, to join organic compounds in a targeted manner, under normal conditions without protecting other amino acid residues. It is possible to introduce protected or modified groupings for chemical transformations. This approach may be relevant primarily in drug design. It is also possible to include amino acids already with modifications to study the influence of posttranslational modifications. Certain ncAAs may have fluorescent groups, or groups that can be registered by other methods, which may be interesting for the lifetime study of biological objects. Another use case is the modification of functional atoms to change affinity or reactivity. The goal of the research group at university Sirius is primarily to incorporate ncAAs for the design of peptides and proteins with unique properties for the creation of pharmaceutical substances.

At present, peptides with ncAAs incorporated by biological synthesis are not widely recognized, whereas ncAAs incorporated by chemical synthesis are widely used. This indicates that this direction is promising.

Methods and Algorithms: The first step in the development of biological systems aimed at obtaining amino acid sequences with ncAA is to obtain aminoacyl tRNA synthetase (aaRS)/tRNA pairs orthogonal to other amino acids, tRNA and aaRS in the cell. aaRS does not interact with other tRNAs and amino acids. tRNA does not interact with other aaRS in the cell. However, tRNA must interact with elements of the ribosome translation system and elongation factors.

In addition to the 20 amino acids that are incorporated into proteins in most organisms, there are two more, selenocysteine and pyrrolysine. Selenocysteine is incorporated using additional mechanisms, while pyrrolizine is incorporated using the traditional mechanism of ribosomal synthesis with the help of an additional aaRS/tRNA pair. This pair was discovered in the archaea genome and was found to be orthogonal to the host aaRS/tRNA pairs of both bacteria and eukaryotes [2]. It was this pair that became a donor for the acquisition of a sufficiently wide variety of ncAAs into proteins and peptides.

The incorporation of more than 200 variants of ncAAs into proteins, following directed modifications, has now been demonstrated. These are mostly lysine and phenylalanine/tyrosine derivatives [3].

Although a wide range of ncAAs have been shown to be incorporated, in most cases the incorporation efficiency is quite low. More work is required to improve both the selectivity of inclusion of specific ncAAs and the efficiency of aaRS. Two basic principles are possible to modify the properties of aaRS. The first principle is rational design. The second principle is directed evolution. We have used the second principal. The PACE (phage assisted directed evolution) method was used in this work [4].

Results: Using PACE, variants capable of incorporating a number of tyrosine derivative amino acids were generated from the pyrrolysine aaRS/tRNA pair: O-allyl-tyrosine, o-triazide-tyrosine, o-methyl-tyrosine, p-amine-phenylalanine, and 2-chloro-phenylalanine. In the first step, directed evolution of aaRS was performed in the presence of 10 mMol of the target amino acids. At lower concentrations, the first step of evolution could not be performed. In the course of successive iterations of directed evolution, it was possible to achieve the system operation when the concentration of target ncAAs was reduced to 2.5 mMol. The works on increasing the efficiency of aaRS, continue.

To analyze the properties of the obtained aaRS, they were cloned into *E. coli* strains expressing the gene of GFP protein with amber stop codon in the 6 position. The performance was evaluated by fluorescence level. Mass spectrometric Analysis of GFP with incorporated o-allyl-tyrosine showed that aaRS variants from the initial stages of PACE have not only lower efficiency but also lower selectivity. That is, they can include the natural amino acids tyrosine and phenylalanine in addition to the target amino acid. aaRS variants from later stages of the PACE experiment had significantly higher specificity.

Conclusion: Strains producing proteins with ncAAs were obtained for the first time in Russia. We have developed a protocol for obtaining orthogonal pairs of aaRS/tRNA to incorporate phenylalanine/tyrosine derivatives into target ncAAs. This protocol can be extended to include lysine, histidine and tryptophan derivatives. Since we can use a single variant, the amber stop codon, it is currently possible to produce peptides and proteins with a single ncAA variant. Expanding the possibilities for parallel incorporation of multiple ncAAs requires the development of multiple, mutually orthogonal pairs and work on other transcription and translation elements.

Funding: The study is supported supported by the grant of the state program of the “Sirius” Federal Territory “Scientific and technological development of the Sirius Federal Territory”.

Список литературы/References

1. Shandell M.A., Tan Z., Cornish V.W. Genetic code expansion: a brief history and perspective. *Biochemistry*. 2021;60(46):3455-3469
2. Wang L., Schultz P.G. A general approach for the generation of orthogonal tRNAs. *Chem Biol.* 2001;8(9):883-890
3. Guo L.-T. et al. Ancestral archaea expanded the genetic code with pyrrolysine. *J Biol Chem.* 2022;298(11):102521
4. Bryson D.I. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat Chem Biol*. 2017;13(12):1253-1260

doi 10.18699/bgrs2024-8.3-05

Исследование специфичности новой метионин аминопептидазы для биотехнологии удаления инициирующего метионина из рекомбинантных белков

Соколов А., Быков В., Вологжанникова А., Трунилина М., Кудряшов Т., Лаптева Ю.*

Федеральный исследовательский центр «Пущинский научный центр биологических исследований» РАН, Институт биологического приборостроения, Пущино, Россия

** yulia.s.lapteva@gmail.com*

Ключевые слова: метионин аминопептидаза; металл-связывающие белки; гетерологичная экспрессия; удаление N-концевого метионина

Мотивация и цель: Отщепление N-концевого инициаторного метионина (iMet) из полипептидной цепи является главной ко- и пост-трансляционной модификацией, охватывающей до 80 % протеома клеток, которая осуществляется специфическими ферментами – метионин аминопептидазами (МАП) [1]. Существенной проблемой в биотехнологии получения рекомбинантных белков, при их гетерологичной экспрессии в *E. coli,* является неполное удаление iMet, что приводит к неправильной укладке белков (агрегации, амилоидоизации) [2], изменению времени их полужизни, потере активности, иммуногенности и др. [3–5]. Для решения этой проблемы применяют технологию обработки целевых рекомбинантных белков ферментами МАП, как в условиях *in vitro*, так и при помощи ко-экспрессии [6–8]. МАП из разных организмов являются металл-зависимыми протеазами и имеют схожую трехмерную укладку, но отличаются субстратной специфичностью и условиями функционирования. Имеющиеся в настоящее время коммерческие МАП человека, *E. coli* и *P. furiosus* не обеспечивают всего разнообразия специфичностей и условий, востребованных в биотехнологии удаления iMet из рекомбинантных белков [9, 10]. В этой связи актуален поиск новых МАП, стабильных в «нестандартных» условиях (повышенных температур, pH, в присутствии денатурантов и др.). Цель данной работы заключалась в исследовании субстратной специфичности новой МАП из бактерии *Thermus thermophilus* (Tth-МАП) с использованием синтетических пептидов.

Методы и алгоритмы: Клонирование и очистку фермента Tth-МАП проводили как описано ранее [11]. Для определения субстратной специфичности Tth-МАП использовали синтетические пептидные субстраты, отличающиеся по аминокислотному остатку после iMet. Реакцию гидролиза пептидов проводили при 60 °C в течение часа в 50 мкл реакционной смеси следующего состава: 50 мМ Hepes pH 7.5, 150 мМ NaCl, 0.1 мМ кобальт, 200 мкМ пептида, 2 мкМ Tth-МАП. Реакцию останавливали добавлением 1 мМ ЭДТА. Степень гидролиза пептидов оценивали при помощи масс-спектрометрии на масс-спектрометрическом детекторе с электрораспылительной ионизацией (ESI-MS) Shimadzu Corporation, LCMS-2010EV (Shimadzu Co., Япония). Калибровку прибора проводили с использованием мио-

глобина сердца лошади (Sigma-Aldrich Co., США). Молекулярные массы белковых фракций анализировали при помощи программного обеспечения GPMAW 9.02 [12].
Результаты: Одной из главных характеристик фермента является его субстратная специфичность. Специфичность МАП варьирует от организма к организму, может меняться в зависимости от связанного катиона металла, но в целом она ограничивается радиусом боковой цепи аминокислотного остатка (а. о.) стоящего за отщепляемым iMet. Обычно это небольшие, незаряженные а. о., такие как Gly, Ala, Ser, Thr, Pro, Val, Cys. В данной работе мы исследовали субстратную специфичность нового фермента МАП из термофильной бактерии *Thermus thermophilus*, а именно провели оценку эффективности отщепления Met из синтетических пептидов. Для стимуляции активности Tth-МАП использовали ионы кобальта (100 мкМ). Анализ показал, что эффективность отщепления Met, за которым находятся Gly и Ala составляет 95 %, в то время как для Thr 15 %. Новый фермент Tth-МАП специфичен к отщеплению Met в пептидах преимущественно с Gly и Ala на N-конце.
Выводы: Впервые изучена субстратная специфичность нового фермента МАП из бактерии *T. thermophilus*. Новый фермент может применяться в биотехнологии для удаления N-концевого метионина из рекомбинантных белков и пептидов.
Финансирование: исследование выполнено при финансовой поддержке Российского научного фонда, грант № 23-24-00563, https://rscf.ru/project/23-24-00563/.

Study of the specificity of a new methionine aminopeptidase for the biotechnology of removing initiating methionine from recombinant proteins

Sokolov A., Bykov V., Vologzhannikova A., Trunilina M., Kudryashov T., Lapteva Yu.*

Federal Research Center "Pushchino Scientific Center for Biological Research" RAS, Institute of Biological Instrumentation, Pushchino, Russia

** yulia.s.lapteva@gmail.com*

Key words: methionine aminopeptidase; metal-binding proteins; heterologous expression; removal of N-terminal methionine

Motivation and Aim: The cleavage of N-terminal initiator methionine (iMet) from the polypeptide chain is the main co- and post-translational modification, covering up to 80 % of the cell proteome, which is carried out by specific enzymes – methionine aminopeptidases (MAP) [1]. A significant problem in the biotechnology of producing recombinant proteins, with their heterologous expression in *E. coli*, is the incomplete removal of iMet, which leads to incorrect protein folding (aggregation, amyloidization) [2], changes in their half-life, loss of activity, immunogenicity, etc. [3–5]. To solve this problem, the technology of processing target recombinant proteins with MAP enzymes is used, both *in vitro* and using co-expression [6–8]. MAPs from different organisms are metal-dependent proteases and have similar three-dimensional folding, but differ in substrate specificity and operating conditions. Currently available commercial MAPs

from humans, *E. coli* and *P. furiosus* do not provide the full variety of specificities and conditions required in the biotechnology of removing iMet from recombinant proteins [9, 10]. In this regard, the search for new MAPs that are stable under "non-standard" conditions (elevated temperatures, pH, in the presence of denaturants, etc.) is relevant. The purpose of this work was to study the substrate specificity of a new MAP from the bacterium *Thermus thermophilus* (Tth-MAP) using synthetic peptides.

Methods and Algorithms: Cloning and purification of the Tth-MAP enzyme was carried out as described previously [11]. To determine the substrate specificity of Tth-MAP, synthetic peptide substrates differing in the amino acid residue after iMet were used. The peptide hydrolysis reaction was carried out at 60 ºC for an hour in 50 µl of a reaction mixture of the following composition: 50 mM Hepes pH 7.5, 150 mM NaCl, 0.1 mM cobalt, 200 µM peptide, 2 µM Tth-MAP. The reaction was stopped by adding 1 mM EDTA. The degree of peptide hydrolysis was assessed using mass spectrometry using an electrospray ionization mass spectrometry detector (ESI-MS) Shimadzu Corporation, LCMS-2010EV (Shimadzu Co., Japan). The device was calibrated using horse heart myoglobin (Sigma-Aldrich Co., USA). The molecular weights of protein fractions were analyzed using GPMAW 9.02 software [12].

Results: One of the main characteristics of an enzyme is its substrate specificity. The specificity of MAP varies from organism to organism and may vary depending on the metal cation bound, but in general it is limited by the side chain radius of the amino acid residue (a. a.) behind the iMet being cleaved off. Usually these are small, uncharged a. a., such as Gly, Ala, Ser, Thr, Pro, Val, Cys. In this work, we investigated the substrate specificity of the new MAP enzyme from the thermophilic bacterium *Thermus thermophilus*, namely, we assessed the efficiency of Met cleavage from synthetic peptides. Cobalt ions (100 µM) were used to stimulate Tth-MAP activity. The analysis showed that the efficiency of elimination of Met, followed by Gly and Ala, is 95 %, while for Thr it is 15 %. The new Tth-MAP enzyme is specific for Met cleavage in peptides predominantly with Gly and Ala at the N-terminus.

Conclusion: The substrate specificity of the new MAP enzyme from the bacterium *T. thermophilus* was studied for the first time. The new enzyme can be used in biotechnology to remove N-terminal methionine from recombinant proteins and peptides.

Funding: research was funded by a grant from the Russian Science Foundation, No. 23-24-00563, https://rscf.ru/project/23-24-00563/.

Список литературы/References

1. Gamerdinger M., Deuerling E. Cotranslational sorting and processing of newly synthesized proteins in eukaryotes. *Trends Biochem Sci*. 2024;49(2):105-118
2. Patke S. et al. Characterization of the oligomerization and aggregation of human Serum Amyloid A. *PLoS One*. 2013;8(6):e64974
3. Liu B. et al. Removal of the N-terminal methionine improves the sweetness of the recombinant expressed sweet-tasting protein brazzein and its mutants in Escherichia coli. *J Food Biochem.* 2020;45(3):e13354
4. Nguyen K.T. et al. N-terminal methionine excision of proteins creates tertiary destabilizing N-degrons of the Arg/N-end rule pathway. *J Biol Chem*. 2019;294(12):4464-4476
5. Arif A., Mohammed K., Nadeem M.S. Biochemical and *in silico* evaluation of recombinant E. coli aminopeptidase and *in vitro* processed human interferon alpha-2b. *Turk J Biol*. 2018;42(3):240-249
6. Nandan A., Nampoothiri K.M. Therapeutic and biotechnological applications of substrate specific microbial aminopeptidases. *Appl Microbiol Biotechnol.* 2020;104(12):5243-52577
7. Liao Y.D. et al. Removal of N-terminal methionine from recombinant proteins by engineered E. coli methionine aminopeptidase. *Protein Sci.* 2004;13(7):1802-1810

8. Hwang D.D. et al. Co-expression of glutathione S-transferase with methionine aminopeptidase: a system of producing enriched N-terminal processed proteins in Escherichia coli. *Biochem J*. 1999;338(Pt. 2):335-342
9. Meng L. et al. Overexpression and divalent metal binding properties of the methionyl aminopeptidase from Pyrococcus furiosus. *Biochemistry*. 2002;41(23):7199-208
10. Roderick S.L., Matthews B.W. Structure of the cobalt-dependent methionine aminopeptidase from Escherichia coli: a new type of proteolytic enzyme. *Biochemistry*. 1993;32(15):3907-3912
11. Lapteva Y.S. et al. Obtaining overstable methionine aminopeptidase for the removal of methionine from recombinant proteins. *J Biomed*. 2023;19(3 E):47-51
12. Peri S., Steen H., Pandey A. GPMAW – a software tool for analyzing proteins and peptides. *Trends Biochem Sci*. 2001;26(11):687-689

doi 10.18699/bgrs2024-8.3-06

Сверхсинтез L-валина в редактированных клетках *Corynebacterium glutamicum* и причины его нестабильности

Шереметьева М.*, Дерунец А., Розанцева В., Леонова Т., Герасимова Т., Яненко А.

НИЦ «Курчатовский институт», Курчатовский геномный центр, Москва, Россия
** m.e.sheremetieva@gmail.com*

Ключевые слова: *Corynebacterium glutamicum*; штамм-продуцент; L-валин; ацетолактатсинтаза

Мотивация и цель: L-валин (далее – валин) – незаменимая аминокислота, участвующая не только в синтезе белков, но также в регуляторных процессах внутри клеток. Валин используется в различных областях производства, например, в фармакологии и косметологии, но наиболее широкое применение находит в качестве кормовой добавки для сельскохозяйственных животных [1].
Основные объёмы валина получают биотехнологическим способом, с помощью штаммов-продуцентов. Популярная основа для таких штаммов – безопасная и неприхотливая почвенная бактерия *Corynebacterium glutamicum* [2]. В нашей стране собственное производство валина пока отсутствует. Современные условия делают необходимым создание такого производства, включая разработку штаммов-продуцентов.
Мы сконструировали штамм *C. glutamicum* VG с уровнем продукции валина в лабораторном ферментёре не менее 70 г/л, что близко к растворимости валина в воде при 30 °C. Однако выяснилось, что продуктивность штамма после нескольких пассажей может опуститься до 40 г/л и ниже (штамм VG-X). Нашей задачей было установить причину этого явления, чтобы предотвратить падение продукции валина при использовании штамма в биотехнологическом процессе.
Методы и алгоритмы: Ферментацию штамма VG и его производных осуществляли в лабораторных ферментёрах на среде с глюкозой и кукурузным экстрактом, а также в пробирках на среде с глюкозой и гидролизатом пшеничного белка. Продуктивность культур оценивали по содержанию L-валина в культуральной жидкости, измеренному с помощью ВЭЖХ. Определение активности ключевого фермента биосинтеза валина ацетолактатсинтазы (AHAS) осуществляли методом Го с соавт. [3]. Секвенирование генов производили по Сэнгеру. Для моделирования белковых молекул применяли программу AlphaFold2 [4]. Для сверхэкспрессии генов использовали автономную плазмиду pNS2 [5].
Результаты: Анализ серии штаммов, полученных из отдельных клонов штамма VG-X, выявил, что эти штаммы имеют разную продуктивность по валину – от 0 до 100 % продуктивности штамма VG. Измерение активности фермента AHAS показало, что она также составляет от 0 до 100 % активности AHAS у штамма VG, причём варьирует пропорционально уровню продукции. Секвенирование генов *ilvBN* в этих штаммах позволило выявить ряд точечных замен нуклеотидов, среди которых – значимые мутации в гене *ilvN*. Моделирование белка IlvN (регуляторная субъединица AHAS) показало, что эти мутации вызывают изменение его структуры, результатом чего может стать нарушение функционирования фермента. Трансформация штамма, несущего одну из таких мутаций, плазмидой

pNS2 с геном *ilvN* без мутации позволила восстановить активность AHAS и продукцию валина. Отсюда следует, что падение продуктивности у этого штамма, действительно, было связано с инактивацией AHAS из-за мутации в гене *ilvN*.
Выводы: Мы предполагаем, что снижение активности AHAS в штамме VG ведёт к перенаправлению потока метаболитов от биосинтеза валина на другие процессы, поэтому клетки с соответствующими мутациями растут быстрее и накапливаются в культуре. Необходимая мера для того, чтобы продуктивность штамма VG при промышленном культивировании оставалась высокой – контроль уровня активности AHAS в посевном материале.
Финансирование: Исследование частично поддержано Министерством науки и высшего образования РФ (грант № 075-15-2019-1659) и Государственным заданием НИЦ «Курчатовский институт».

L-valine supersynthesis in edited *Corynebacterium glutamicum* cells and the reasons for its instability

Sheremetieva M.*, Derunets A., Rozantseva V., Leonova T., Gerasimova T., Yanenko A.
NRC "Kurchatov Institute", Kurchatov Genomic Center, Moscow, Russia
** m.e.sheremetieva@gmail.com*

Key words: *Corynebacterium glutamicum*; producer strain; L-valine; acetolactate synthase

Motivation and Aim: L-valine (hereinafter referred to as valine) is an essential amino acid involved not only in protein synthesis but also in regulatory processes within cells. Valine is used in various fields of production, for example, in pharmacology and cosmetology, but it is most widely used as a feed additive for farm animals [1].
The main amounts of valine are obtained biotechnologically, using strains-producers. A popular basis for such strains is the safe and unpretentious soil bacterium *Corynebacterium glutamicum* [2]. In Russia, domestic production of valine still doesn't exist. The current situation dictates the necessity to create such production, including the development of strains-producers.
We designed a strain of *C. glutamicum* VG with a level of valine production in the laboratory fermenter of at least 70 g/L, which is close to the solubility of valine in water at 30 °C. However, it was found that the productivity of the strain can drop to 40 g/L or lower after several passages (VG-X strain). Our goal was to determine the cause of this phenomenon in order to prevent a drop in valine production when the strain is used in a biotechnological process.
Methods and Algorithms: Fermentation of VG strain and its derivatives was carried out in laboratory fermenters on a medium with glucose and a corn extract, and in test tubes on a medium with glucose and wheat protein hydrolysate. The productivity of the cultures was assessed by the L-valine content in the culture fluid measured by HPLC. The activity of the key enzyme of valine biosynthesis, acetolactate synthase (AHAS), was determined by the method of Guo et al. [3]. Gene sequencing was performed by Sanger sequencing. AlphaFold2 program [4] was used for modeling protein molecules. The autonomous plasmid pNS2 was used for gene overexpression [5].

Results: The analysis of a series of strains derived from individual clones of VG-X strain revealed that these strains have different valine productivity, ranging from 0 to 100 % of the productivity of VG strain. Measurement of AHAS enzyme activity showed that it also ranged from 0 to 100 % of AHAS activity in VG strain, with variation proportional to the level of production. Sequencing of the *ilvBN* genes in these strains revealed a number of point nucleotide substitutions, including significant mutations in the *ilvN* gene. Modeling of the IlvN protein (regulatory subunit of AHAS) showed that these mutations cause changes in its structure, which may result in impaired functioning of the enzyme. Transformation of a strain carrying one of these mutations with plasmid pNS2 with the *ilvN* gene without the mutation allowed to restore AHAS activity and valine production. It follows that the drop in productivity in this strain was indeed due to the inactivation of AHAS due to the mutation in the *ilvN* gene.
Conclusion: We assume that a decrease in AHAS activity in VG strain leads to a redirection of metabolite flux from valine biosynthesis to other processes, so cells with the corresponding mutations grow faster and accumulate in culture. A necessary measure to ensure that the productivity of VG strain in industrial cultivation remains high is to control the level of AHAS activity in the inoculum.
Funding: This research was partially funded by the state task of the National Research Center “Kurchatov Institute” and by the Ministry of Science and Higher Education of the Russian Federation № 075-15-2019-1659.

Список литературы/References

1. Hao Y. et al. Microbial production of branched chain amino acids: Advances and perspectives. *Bioresour Technol.* 2024;397:130502. doi 10.1016/j.biortech.2024.130502
2. Sheremetieva M.E. et al. Rational metabolic engineering of Corynebacterium glutamicum to create a producer of L-valine. *Vavilovskii Zhurnal Genet Selektsii*. 2022;26(8):743-757. doi 10.18699/VJGB-22-90
3. Guo Y. et al. Analysis of acetohydroxyacid synthase variants from branched-chain amino acid-producing strains and their effects on the synthesis of branched-chain amino-acids *in Corynebacterium glutamicum*. *Protein Expr Purif.* 2015;109:106-112. doi 10.1016/j.pep.2015.02.006
4. Mirdita M. et al. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022;19:679-682. doi 10.1038/s41592-022-01488-1
5. Ростова Ю.Г. и др. Конструирование челночного вектора для коринебактерий и *E. coli*. Клонирование и изучение экспрессии гена *lysC*. *Биотехнология*. 1993;4:14-17
[Rostova Y.G. et al. Construction of novel shuttle cloning vector for Corynebacteria and *E. coli*. En Cloning and expression of *lysC* gene. *Biotekhnologiya*. 1993;4:14-17 (in Russian)]

doi 10.18699/bgrs2024-8.3-07

Genetic technologies: production of fermentative preparations for industrial needs based on the construction of producers using a collection of natural genes of bacteria and the genome of the yeast *Komagataella phaffii*

Peltek S.E.[1, 2]*, Zadorozhny A.V.[1, 2], Bogacheva N.V.[1, 2], Korzhuk A.V.[1, 2], Shlyakhtun V.N.[1, 2], Uvarova Y.E.[1, 2], Shipova A.A.[1, 2], Arbuzov G.D.[1, 2], Chesnokov D.O.[1, 2], Bannikova S.V.[1, 2], Pavlova E.Yu.[1, 2], Bochkov D.V.[1, 2], Bukatich E.Yu.[1, 2], Slyunko N.M.[1, 2], Bryanskaya A.V.[1, 2], Khlebodarova T.M.[1, 2], Vasilieva A.R.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *peltek@bionet.nsc.ru*

Key words: strains-producers; enzymes; cultural liquids; proteomes

Motivation and Aim: The main objective of this work is to create a conveyor for constructing strains-producers of enzymes for industry, which will enable the import substitution of essential enzymes for industrial needs and the development of new enzyme (protein) preparations absent in global markets "on a turnkey basis".

Methods and Algorithms: This task is partially implemented based on the Collection of microorganisms for biotechnological purposes of the Federal Research Center Institute of Cytology and Genetics SB RAS, more than 2,100 genomes of natural strains of which have been sequenced and annotated to date. A database of genes controlling the biosynthesis of enzymes used in various biotechnological productions has been formed. Experimental studies of the main properties of enzymes expressed by annotated genes have been conducted (substrate specificity, thermostability, optimum pH, activities, pI points, etc.). The nucleotide sequences of the studied enzymes were used for directed gene-specific obtaining of strains-producers of necessary enzymes.

Results: Using genetic engineering methods, their transfer into the genome of the yeast *Komagataella* phaffii has been performed, effective expression of target proteins in the cultural medium under the control of the AOX1 gene promoter and terminator and the leader peptide has been shown, and a technology for obtaining preparations of target proteins (enzymes) has been developed, including methods of cross-flow membrane filtration. During the cultivation of the strain T07 based on *K. phaffii* cells, a producer of alpha-amylase from *Bacillus licheniformis* 47018, studies were conducted to determine the target enzyme with confirmation of its structure and quantitative comparison of its content. of this enzyme in cultural liquids and proteomes. The created basic methods of qualitative and quantitative analysis of molecules of target enzymes make it possible to reliably determine their presence in cultural liquids and preparations.

Conclusion: The efficiency of the obtained conveyor has been tested during semi-industrial synthesis of alpha-amylase from *Bacillus licheniformis* 47018 under the conditions of a 100-liter bioreactor.

Funding: The work was carried out with the financial support of the "Kurchatov Genomic Center of the Federal Research Center Institute of Cytology and Genetics SB RAS" (No. 075-15-2019-1662) and the Ministry of Science and Higher Education of the Russian Federation (No. FWNR-2022-0022).

doi 10.18699/bgrs2024-8.3-08

Genomic analysis of adaptive laboratory evolution of *Rhodococcus rhodochrous* strains – biocatalysts for acrylic monomers production

Shemyakina A.O.*, Grechishnikova E.G., Novikov A.D., Lavrov K.V., Yanenko A.S.
Kurchatov Genomic centre, National Research Centre "Kurchatov Institute", Moscow, Russia
** shemyakina.a.o@gmail.com*

Key words: adaptive laboratory evolution; *Rhodococcus*; nitrile hydratase; amidase

Motivation and Aim: A series of improved *Rhodococcus rhodochrous* biocatalyst strains for bioproduction of industrially significant acrylic monomers (acrylamide, acrylic acid, N-substituted acrylamides) was developed earlier in our laboratory using adaptive laboratory evolution (ALE) and further precise genome editing [1]. The ALE approach was chosen based on the hypothesis of co-regulation of nitrile hydratase (NHase) and amidase genes, conferring the corresponding activities, required for monomer production. The aim of this work was to track changes, appeared in the aforesaid genes, as well as in regulatory genes, which are assumed to control their expression. Additionally, we aimed to track whole-genome changes occurred during the ALE experiment, in order to identify the mutation types responsible for evolutionary plasticity of *R. rhodochrous* genome.
Methods and Algorithms: the genomes of *Rhodococcus rhodochrous* strains and their annotations [2] are available in the NCBI database under accession numbers NZ_CP064065.1, NZ_CP064066.1 (M8 strain genome and megaplasmid). NZ_CP129464.1, NZ_CP129465.1, NZ_CP129463.1 (M8-35 strain genome, 0.17 Mbp megaplasmid and 0.155 Mbp megaplasmid), NZ_CP129391.1, NZ_CP129392.1 (M8-20 strain genome and megaplasmid), NZ_CP129335.1, NZ_CP129336.1 (M8-33 strain genome and megaplasmid), NZ_CP129333.1, NZ_CP129334.1 (M8-50 strain genome and megaplasmid). Short changes (SNPs, indels) between the genomes were identified using Snippy [3] with default settings and --contig option. Larger changes (rearrangements) were identified using progressive Mauve algorithm [4]. Differences in amino acid sequences of enzymes associated with target activities were additionally verified using BLAST [5] algorithm supported by custom Python scripts.
Results: The wild-type *Rhodococcus rhodochrous* M8 strain which genome consists of a chromosome (6.1 Mbp) and a linear megaplasmid (0.17 Mbp) [6] served as the starting point of ALE experiment. Nitrile hydratase activity of M8 strain is determined by *nhmBA* genes, coding NHase β and α subunits [7]. Identification of genes responsible for amidase activity was complicated by the abundance of amidase genes (32) in M8 genome. The ALE experiment that aimed to exploit amidase phenotype of M8 strain consisted of 2 stages: (*i*) selection of clones with suppressed target activities and capable of growth in the presence of the toxic chloroacetamide (representative strain: M8-35) and (*ii*) obtaining revertant strains selected on medium that contained acetamide as the sole carbon source (representative strains: M8-20, M8-33, M8-50).
The results of tracking of changes in genes relative to target activities are summarized in Table 1. M8-35 was compared to wild-type genome and then served as a reference for

identifying mutations in genomes of revertant strains. Aside from enzyme genes mutations were concentrated in genes *nhmCD* of amide-dependent NHase transcription regulators. Their close homologues *nhlC and nhlD* are responsible for positive and negative regulation of corresponding NHase genes, respectively [8].

Table 1. Changes in genes relative to NHase and amidase activities in *R. rhodochrous* M strains series obtained in ALE experiment and their corresponding phenotypes

Gene name	Location	Changes occurred in				Product
		M8-35	M8-20	M8-33	M8-50	
amiE	Mega-plasmid	NC*	NC	Deletion	NC	Aliphatic amidase
fmdA1	Mega-plasmid	NC	NC	Deletion	NC	Acetamidase/formamidase
nhmC	Mega-plasmid	IS integration	NC	Deletion	Deletion	Amide-dependent NHase transcription regulator
nhmD	Mega-plasmid	NC	IS integration	Partial deletion	Deletion	Amide-dependent NHase transcription regulator
nhmB	Mega-plasmid	NC	NC	NC	Deletion	NHase subunit beta
nhmA	Mega-plasmid	NC	NC	NC	Deletion	NHase subunit alpha
amiN1	Chromosome	two FS**	NC	NC	NC	Amidohydrolase
amiN2	Chromosome	NC	NC	NC	FS	Amidohydrolase
Phenotype:	NHase	< 0.7	109.4	114.5	< 0.7	
	Amidase	< 1.5	106	12.4	210.4	

* NC – no changes. ** FS – frameshift mutation. Phenotype is expressed as percent of initial activity level of M8 strain.

Overall genome alignment revealed comparable scale of mutation background in all pairs of parent-child strains: about 200 point mutations, including SNPs (9–24 %), indels (74–90 %) and their combinations, and about 9 large (8 bp – 52 kbp long) rearrangements. Depending on the genomes being compared 24–62 % of point mutations led to changes in amino acid sequence of coding genes. The repertoire of large rearrangements included: IS relocations, tandem repeat (TR) contractions and expansions, gene duplication, 10s kbp long deletions and prophage integration; in half the cases the mutations were intragenic. Additionally, 48 % of large rearrangements that occurred during the second ALE stage were completely reversable, including recovery of ATP-binding protein gene (gen tag ISO16_RS07605) in all three revertant strains after IS mediated knockout in M8-35 strain. Besides IS relocations reversable cases belonged to TR copy number variation and the duplication of tRNA gene (Fig. 1). Aside from prophage integration another intercellular transposition was the appearance of second 0.155 Mbp long megaplasmid in M8-35 genome which however was not found in any of analyzed genomes of M8-35 offsprings.

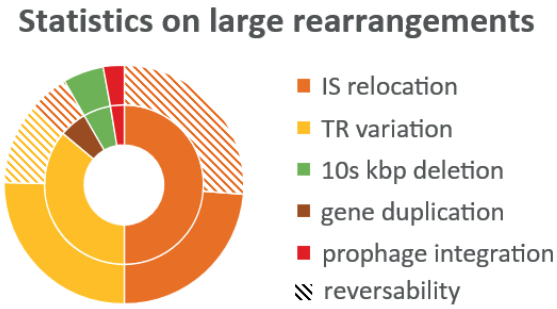


Fig. 1. Distribution of large rearrangements by type and reversibility in *R. rhodochrous* M strains series, obtained in ALE experiment

Conclusion: The two types of genome modifications that targeted the genes critical for NHase and amidase activities in M strains were gene disruption by IS integration and large deletions of unknown origin. In contrast, point mutations were much less associated with changes in key genes despite the considerable amount of indel induced frameshift mutations. Notably, most of key genes are located on the megaplasmid, which may indicate its evolutionary role in the metabolic adaptation of *Rhodococcus* bacterium to such recalcitrant substrates as nitriles.

In M8-33 and M8-50 strains the suppressed enzymatic activities conspicuously matched with deletions of corresponding enzymatic genes. Meanwhile the simultaneous switching of both activities in a row of M8 –> M8-35 –> M8-20 strains was accompanied by subsequent knockouts of the *nhmC* and *nhmD* genes. The fact that both knockouts were IS mediated leaves the possibility of full gene restoration as it was shown for another gene in this ALE set. Together these observations allow to assume the major role of *nhmCD* genes in nitrile-amide metabolism regulation in M8 strain.

Funding: The study is supported by the State Assignment for the National Research Center Kurchatov Institute.

References

1. Lavrov K.V., Shemyakina A.O., Grechishnikova E.G. et al. A new concept of biocatalytic synthesis of acrylic monomers for obtaining water-soluble acrylic heteropolymers. *Metab Eng Commun.* 2024;18:e00231. doi 10.1016/j.mec.2023.e00231
2. Tatusova T., DiCuccio M., Badretdin A. et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016;44(14):6614-6624. doi 10.1093/nar/gkw569
3. Seemann T. Rapid haploid variant calling and core genome alignment. 2015. https://github.com/tseemann/snippy
4. Darling A.E., Mau B., Perna N.T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS One.* 2010;5(6):e11147. doi 10.1371/journal.pone.0011147
5. Camacho C., Coulouris G., Avagyan V. et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421. doi 10.1186/1471-2105-10-421
6. Novikov A.D., Lavrov K.V., Kasianov A.S., Topchiy M.A., Gerasimova T.V., Yanenko A.S. Complete genome sequence of *Rhodococcus* sp. strain M8, a platform strain for acrylic monomer production. *Microbiol Resour Announce.* 2021;10(10):e01314-20. doi 10.1128/mra.01314-20
7. Lavrov K.V., Shemyakina A.O., Grechishnikova E.G. et al. *In vivo* metal selectivity of metal-dependent biosynthesis of cobalt-type nitrile hydratase in *Rhodococcus* bacteria: a new look at the nitrile hydratase maturation mechanism? *Metallomics.* 2019;11(6):1162-1171. doi 10.1039/c8mt00129d
8. Komeda H., Kobayashi M., Shimizu S. A novel gene cluster including the *Rhodococcus rhodochrous* J1 *nhlBA* genes encoding a low molecular mass nitrile hydratase (L-NHase) induced by its reaction product. *J Biol Chem.* 1996;271(26):15796-15802. doi 10.1074/jbc.271.26.15796

doi 10.18699/bgrs2024-8.3-09

How RNA free energy landscape of *ilvBNC* operon transcriptional attenuator drives *Corynebacterium glutamicum* valine production

Titov I.I.[1, 2]*, Ryabchenko L.E.[3], Leonova T.E.[3], Kalinina T.I.[3], Gerasimova T.V.[3], Khlebodarova T.M.[1, 2], Sheremetieva M.E.[3], Kolchanov N.A.[1, 2], Yanenko A.S.[3]

[1] *Department of Systems Biology, Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *National Research Center "Kurchatov Institute", Kurchatov Genome Center, Moscow, Russia*

* titov@bionet.nsc.ru

Key words: *Corynebacterium glutamicum*; transcription regulation; attenuation; *ilvBNC* operon; acetohydroxy acid synthase AHAS

Motivation and Aim: The adaptation of bacteria to environmental changes has led to their ability to rapidly modulate gene expression in various ways, including RNA-dependent regulation. RNA free energy landscapes translate environmental signals into the language of RNA structural dynamics, which, due to the high specificity and speed of complementary interactions, ensures an immediate reaction of the bacterium. An example is an attenuator – a special RNA element, which frequently regulates bacterial transcription, in particular, of the *ilvBNC* operon of *C. glutamicum*, which then determines the biosynthesis of branched chain amino acids (L-isoleucine, L-leucine, L-valine). The goal of this work was to investigate the mechanism of transcriptional regulation of *ilvBNC* operon expression both experimentally and computationally.

Methods and Algorithms: The *C. glutamicum* strains were received from the Russian Collection Industrial Microorganisms (VKPM) of the NRC "Kurchatov Institute", the German Collection of Microorganisms and Cell Cultures (DSMZ) and Derbikov D. (NCR "Kurchatov Institute", Kurchatov Genome Center).

C. glutamicum strains with the mutations in the *ilvBNC* operon were constructed by replacing the native sequence of the regulatory region with a mutant copy using homologous recombination. The transcription level of the *ilvBNC* genes in the mutated strains was studied using real-time PCR with reverse transcription. The activity of AHAS in the cell-free extracts of strains was measured by the method [1].

We used the Vienna [2] and GArna [3] programs to calculate the secondary structure of the attenuator region of the *ilvBNC* operon and the thermodynamic parameters of [4, 5] to calculate the mutational perturbations in the hairpins free energies.

Results: Fig. 1A shows the organization of the *ilvBNC* operon and the structure of its regulatory region located upstream of the *ilvB* gene [6]. An interesting feature of the attenuator is the overlapping of Ter- and AT-hairpins which allows toehold formation and continuous transition between termination and antitermination states (Fig. 1B).

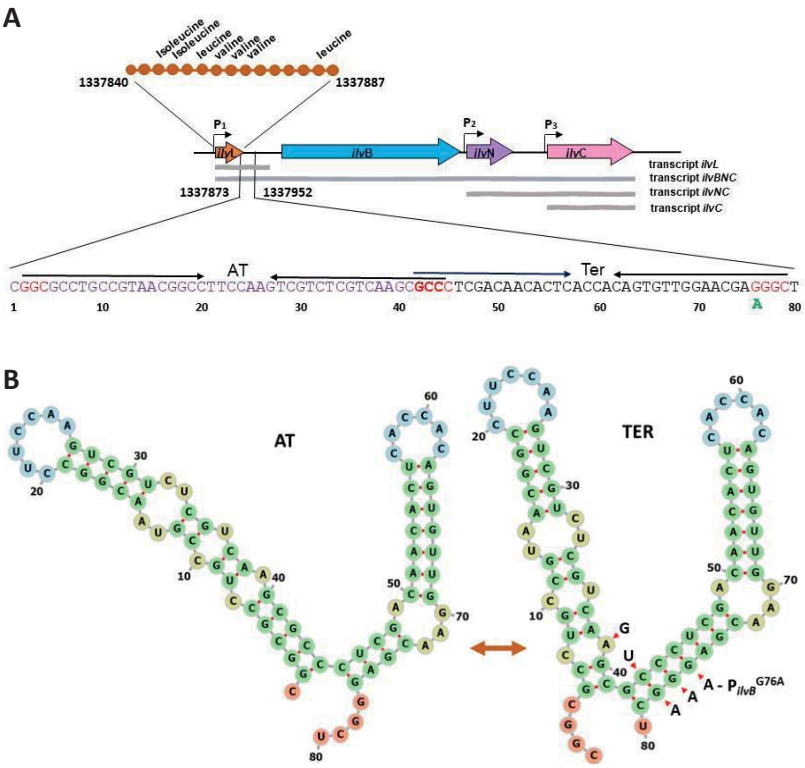


Fig. 1. Organization of the *ilvBNC* operon in *C. glutamicum* and alternative secondary structures of the transcription attenuator in the regulatory region of the *ilvBNC* operon of the *C. glutamicum*.
(A) The location of the sequences of the leader peptide IlvL, the structural genes *ilvB*, *ilvN* and *ilvC*, as well as the structure of the leader peptide and regulatory sequence upstream of the *ilvBNC* operon 80 bp long. Designations: P1, P2, P3 – promoters; the G76A mutation found during whole-genome sequencing of the valine-producing strain *C. glutamicum* VF is marked in green; the sequence GCCC (at positions 42–45) and its complementary GGGC (76–79) and GGC (2–4) are marked in red; the boundary trinucleotide GCC is marked in bold. (B) Alternative secondary structures of the transcription attenuator in the regulatory region of the *ilvBNC* operon of the *C. glutamicum*. Secondary structures responsible for free transcription (AT) and its arrest (TER) are shown on the left and right, respectively. Red arrows indicate mutations introduced into the attenuator structure

We created the strains with the mutations in the attenuator region and measured their characteristics (Fig. 2). A high correlation ($r^2 = 0.9443$, $p < 10^{-3}$) was observed between the level of valine production and the level of *ilvB* gene expression (Fig. 2C), indicating a key contribution of *ilvB* gene transcription to the control of valine production in *C. glutamicum*. The levels of *ilvB* gene expression and valine production also correlated well with the level of AHAS enzyme activity, $r^2 = 0.8926$ and $r^2 = 0.9421$, respectively. A significant correlation ($r^2 = 0.8027$, $p < 0.05$) was also observed between the expression of the *ilvBNC* operon in the *C. glutamicum* mutant strains and the difference in antitermination-termination states free energy according to the structural model (Fig. 1B).

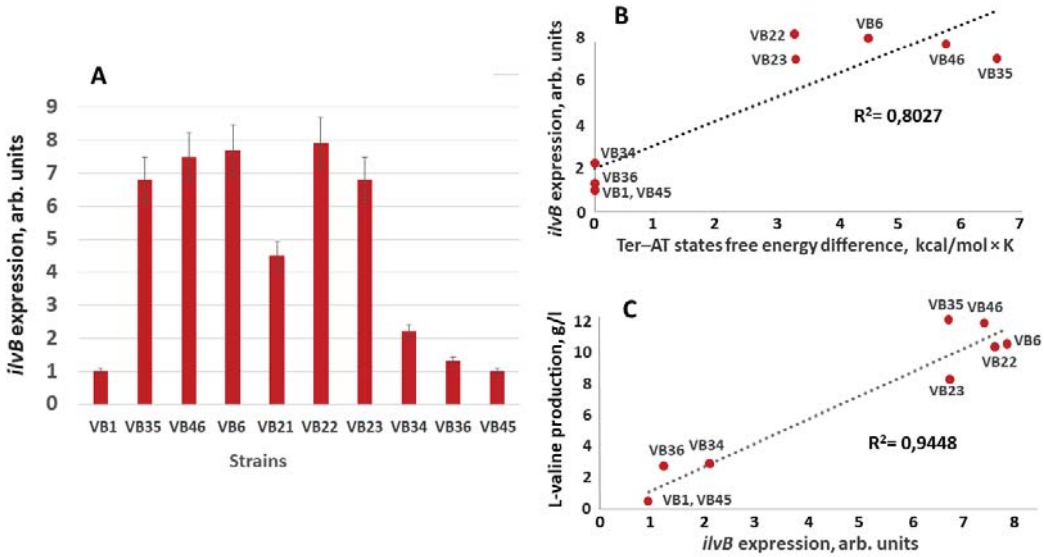


Fig. 2. Relative expression level of the *ilvBNC* operon in the *C. glutamicum* strains with mutations in the Ter-AT region (A), its dependence on changes in the free energy difference of Ter- and AT-hairpins (B), and the dependence of the level of valine production on the relative expression level of the *ilvB* gene in the mutant strains (C). The mutant strains are (see Fig. 1B): VB1 – ATCC13032, VB6 – G76A, VB22 – G76T, VB23 – G76C, VB34 – A39G, VB35 – G77A, VB36 – C43T, VB45 – C43T and G78A, VB46 – G78A

Conclusion: We develop a model of transcriptional regulation of *ilvBNC* operon expression, which explains our experimental data on single-nucleotide substitutions in the regulatory region. In this model, the hairpins of the terminator and the antiterminator partially overlap, so that the boundary trinucleotide GCC participates in the formation of both structures. As a result, in the free energy landscape, a valley is formed between the termination and antitermination states, and the attenuator is able to cross it using a sequence of low-energy transitions, thus controlling rapid changes in valine production.
Funding: The work was supported by the Ministry of Science and Higher Education of the Russian Federation (project No. 075-15-2019-1659) by the Program of Fundamental Studies of the Siberian Branch of the Russian Academy of Sciences (project No. FWNR-2022-0020).

References

1. Guo Y., Han M., Xu J., Zhang W. Analysis of acetohydroxyacid synthase variants from branched-chain amino acids-producing strains and their effects on the synthesis of branched-chain amino acids in Corynebacterium glutamicum. *Protein Expr Purif*. 2015;109:106-112. doi 471 10.1016/j.pep..02.006
2. Lorenz R., Bernhart S.H., Höner zu Siederdissen C., Tafer H., Flamm C., Stadler P.F., Hofacker I.L. ViennaRNA Package 2.0. *Algorithms Mol Biol*. 2011;6:26. 485 doi 10.1186/1748-7188-6-26
3. Titov I., Vorobiev D., Ivanisenko V., Kolchanov N. A fast genetic algorithm for RNA secondary structure analysis. *Russian Chemical Bulletin*. 2002;51:1135-1144
4. Davis A.R., Znosko B.M. Thermodynamic characterization of single mismatches found in naturally occurring RNA. *Biochem*. 2007;46:13425-13436. doi 10.1021/bi701311c
5. Zuber J., Schroeder S.J., Sun H., Turner D.H., Mathews D.H. Nearest neighbor rules for RNA helix folding thermodynamics: improved end effects. *Nucleic Acids Res*. 2022;50:5251-5262. doi 10.1093/nar/gkac261
6. Morbach S., Junger C., Sahm H., Eggeling L. Attenuation control of ilvBNC in Corynebacterium glutamicum: evidence of leader peptide formation without the presence of a ribosome binding site. *J Biosci Bioeng*. 2000;90, 501-507. doi 10.1016/s1389-1723(01)80030-491x

doi 10.18699/bgrs2024-8.3-10

Identification and evaluation of the expression and secretion levels of the yeast genome-derived alpha-amylase enzyme *Komagataella phaffii* by HPLC-MS method

Vasilieva A.R.[1, 2]*, Korzhuk A.V.[1, 2], Zadorozhny A.V.[1, 2], Uvarova Yu.E.[1, 2], Shlyakhtun V.N.[1, 2], Bukatich E.Yu. [1], Chesnokov D.O.[1], Pavlova E.Yu.[1], Peltek S.E.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *vasilieva@bionet.nsc.ru*

Key words: *K. phaffii* T07; HPLC-MS; alpha-amylase; proteome; secretome

Motivation and Aim: Currently, enzyme preparations that hydrolyze starch, which constitutes a significant proportion of carbohydrates in the feed weight but has a low percentage of digestibility, are in demand in agriculture to increase the nutritional value of feed. Alpha-amylase (α-1,4-glucan-4-glucanohydrolase) is an endoamylase and catalyzes the hydrolysis of α-1,4-glycosidic bonds within the chain of starch and related carbohydrates to form substances with a low degree of polymerization such as glucose, maltodextrin, and oligosaccharides of various lengths. Research in the field of α-amylase production is related to the search for strains of highly efficient thermostable enzymes, improvement of various characteristics of synthesis, secretion of recombinant proteins, protection against protease degradation, etc., as well as the search for inexpensive substrates for industrial cultivation of strain-producers. α-amylases from bacteria of *Bacillus subtilis*, *B. amyloliquefaciens*, *B. licheniformis*, and *B. stearothermophilus* species are of industrial importance. Expression systems based on bacteria including *B. subtilis* and *Brevibacillus choshinensis*, the fungus Aspergillus oryzae and the yeast *Pichia pastoris (Komagataella phaffii)* have been used to produce recombinant α-amylases. The aim of the present work was the qualitative identification and relative quantification of the alpha-amylase enzyme derived from the genome of the yeast *Komagataella phaffii*.

Methods and Algorithms: During the study, proteins were purified by quantitative precipitation and concentration followed by tryptic hydrolysis. The prepared and purified peptide mixture was separated using a Thermo Fisher Scientific Ultimate 3000 Series HPLC system (Nano/CapSystemNCS-3500RS) followed by tandem mass spectrometric analysis on an Orbitrap Fusion Lumos. The primary data were processed using the Proteome Discoverer 2.4 software package to obtain highly accurate information on the composition and ratio of proteins in the samples

Results: In samples of recombinant strain *K. phaffii* T07, peptide sequence fragments identifying the structure of α-amylase with the indicated putative amino acid sequence were detected with 88.09 % convergence (Fig. 1). Quantitative ratios of α-amylase enzyme in 4 clones of K. phaffii T07 were obtained and are shown in Fig. 2.

In addition, a comparison of α-amylase secretion in the original α-amylase-producing strain *B. licheniformis* 47018 and secretion of the target α-amylase enzyme into the culture fluid of the α-amylase-producing strain *K. licheniformis 47018* and the secretion of the target enzyme α-amylase into the culture fluid of the α-amylase-producing strain

K. phaffii T07 (Fig. 3, 4). The ratio of amounts (in culture fluid/proteome) of secreted enzyme for *K. phaffii* T07 clones: clone 10 – 220.7/7.1; clone 45 – 82.4/6.9; clone 5 – 209.7/7.1; clone 7 – 245.7/20.4 (normalization was performed for clone 45).

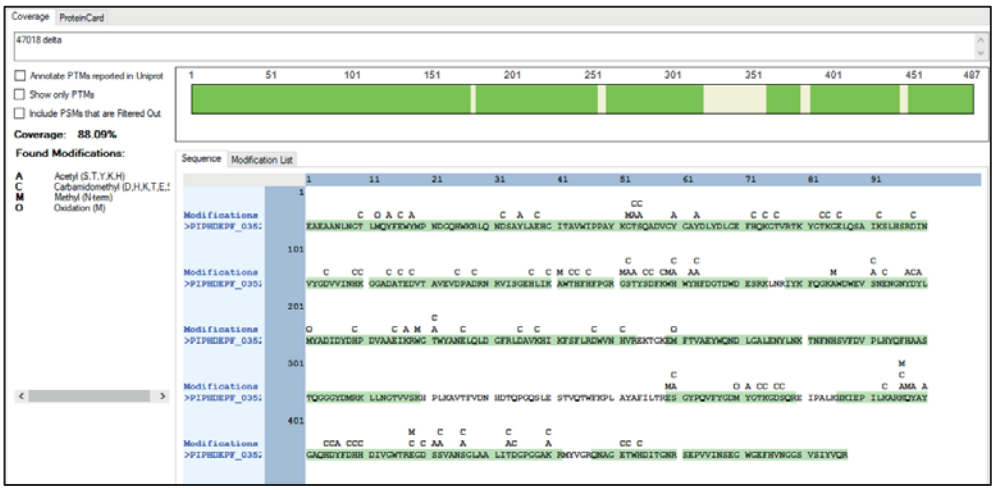


Fig. 1. Identification of the structure of α-amylase with the putative amino acid sequence of α-amylase from *B. licheniformis* 47018

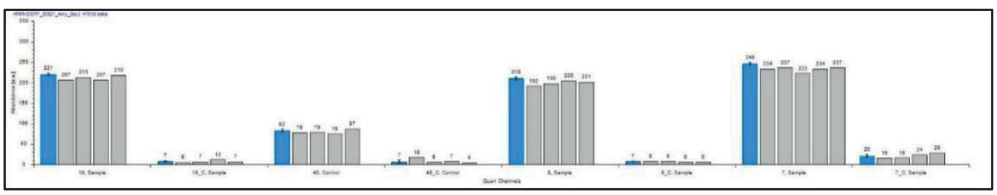


Fig. 2. Quantification of α-amylase enzyme in the proteome (Control) and secretome (Sample) of *K. phaffii* T07 (clones 10, 5, 7) and in the proteome (Control) and secretome (Sample) of *K. phaffii* T07 (clone 45)

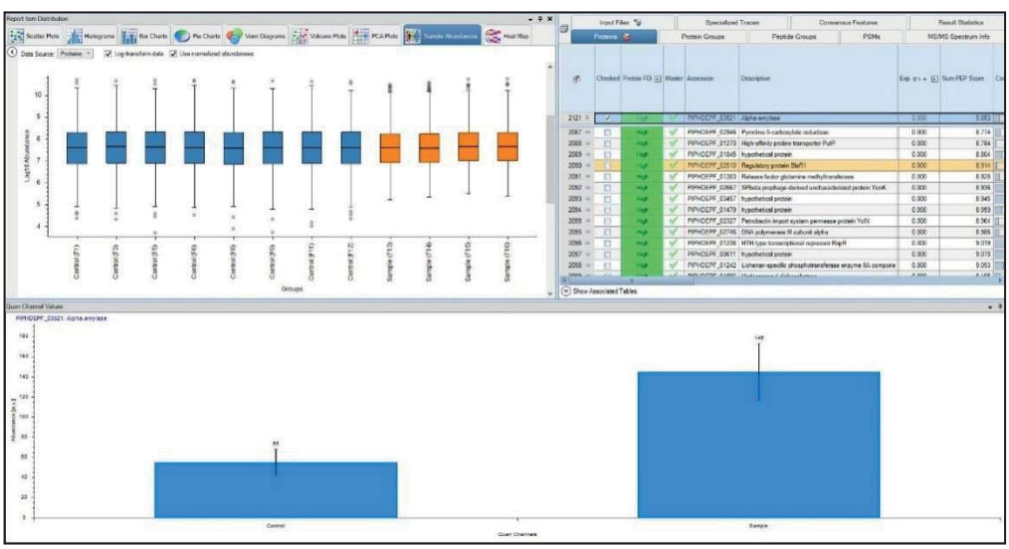


Fig. 3. Identification and quantitative comparison of α-amylase enzyme in the proteome (Control) and secretome (Sample) of *B. licheniformis* 47018

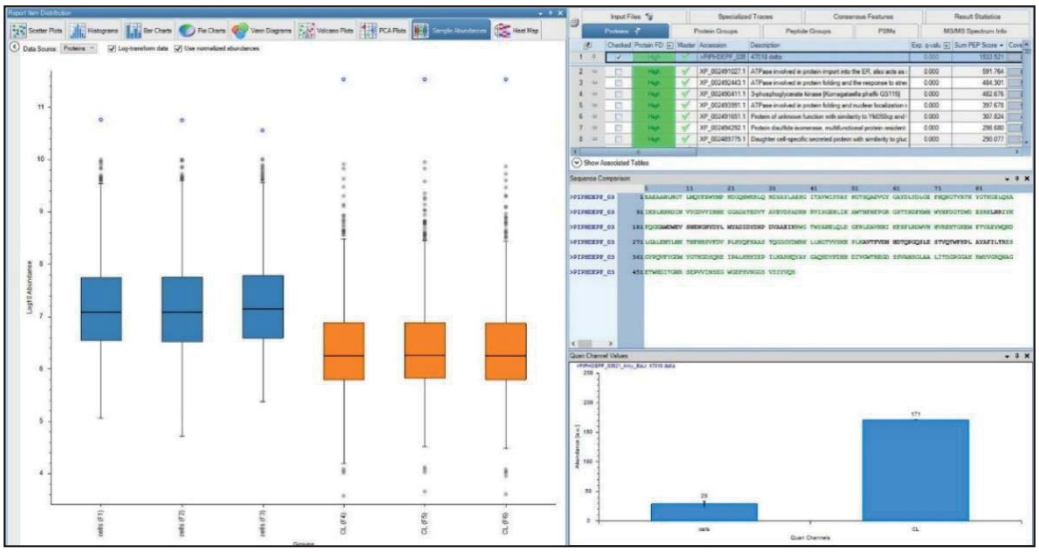


Fig. 4. Identification and quantitative comparison of α-amylase enzyme in the proteome (cells) and secretome (CL) of *Komagataella phaffii* T07

Conclusion: The established basic methods for qualitative and quantitative analysis of molecules of the target enzyme α-amylase from *B. licheniformis* 47018 allow to reliably determine their presence in culture fluids and preparations based on cells of *Komagataella phaffii* strain T07.

Funding: The study was financially supported by the Ministry of Science and Higher Education of the Russian Federation under grant No. FVNR-2022-0022 at the Federal Research Center ICIG SB RAS and under the ICIG SB RAS Core Facility of Proteomic and metabolomic profiling of microorganisms (supported by project No. 075-15-2019-1662).

8 Симпозиум «Микробиология и биотехнологии: компьютерные и экспериментальные подходы»

Symposium “Microbiology and biotechnologies: computational and experimental approaches”

doi 10.18699/bgrs2024-8.4-02

Предсказание количества белка в дрожжах, основанное на трансферном обучении

Вензель А.С.[1, 2]*, Клименко А.И.[1, 2], Деменков П.С.[1, 2, 3], Иванисенко Т.В.[1, 2, 3], Лашин С.А.[1, 2, 3], Колчанов Н.А.[1, 2, 3], Иванисенко В.А.[1, 2, 3]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

[3] *Новосибирский государственный университет, Новосибирск, Россия*

** venzel@bionet.nsc.ru*

Ключевые слова: дрожжи; количество белка; трансформеры; машинное обучение

Прогнозирования уровня белков в бактериях имеет большое значение для биотехнологии и оптимизации процессов биосинтеза, в частности для компьютерного дизайна штаммов-продуцентов с повышенной экспрессией целевых наборов белков. Использование современных подходов машинного обучения улучшает понимание количественной динамики белков и ускоряет разработку новых биотехнологических приложений. И существующие современные методы глубокого обучения, такие как нейросетевые модели обработки естественного языка с архитектурой Transformer, обученные на больших массивах данных, позволяют анализировать биологические последовательности без «ручного» извлечения признаков последовательности, что значительно повышает точность предсказаний моделей машинного обучения. В рамках данной работы был разработан метод предсказаний количества белка в *Saccharomyces cerevisiae* S288C, названный yeastProtPred, включающий в себя совместное использованеи аминокислотных и геномных трансформеров для получения векторных представлений последовательностей гена, а также последовательностей кодируемого белка.

Информация о полном геноме дрожжей Saccharomyces cerevisae S288C и аннотация этого генома были взяты из базы данных The Saccharomyces Genome Database (SGD). Количественные данные по уровням представленности белков в клетках *Saccharomyces cerevisiae* S288C были взяты из Protein Abundance Database (PaxDb). Для получения векторных представлений последовательностей генов и белков использовались предобученные BERT-образные трансформеры GENA-LM и ESM-2 соответственно. Векторные представления используются, как входные параметры для обучения модели градиентного бустинга для предсказания количества белка.

Финансирование: Исследование проведено при финансовой поддержке проекта Министерства науки и высшего образования РФ «Курчатовский центр геномных исследований мирового уровня» № 075-15-2019-1662 от 31.10.2019.

Transfer learning based yeast protein abundance prediction

Venzel A.S.[1, 2]*, Klimenko A.I.[1, 2], Demenkov P.S.[1, 2, 3], Ivanisenko T.V.[1, 2],
Lashin S.A.[1, 2, 3], Kolchanov N.A.[1, 2, 3], Ivanisenko V.A.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *venzel@bionet.nsc.ru*

Key words: yeast; protein abundance; transformer; machine learning

Protein abundance prediction in bacteria is crucial for biotechnology and optimizing biosynthesis processes, particularly for computer-aided producing strains design with enhanced target gene expressions. Modern machine learning approaches enhance our understanding of protein quantitative dynamics and accelerate the development of new biotechnological applications. Existing deep learning methods, such as natural language processing neural network models with Transformer architecture trained on large datasets, allow to analyze biological sequences without manual feature extraction, significantly improving the accuracy of machine learning model predictions.

In this study, a method named yeastProtPred was developed for predicting protein abundance in Saccharomyces cerevisiae S288C. It incorporates the joint use of amino acid and genomic transformers to obtain vector representations of gene sequences and encoded protein sequences.

Information about the complete genome of Saccharomyces cerevisiae S288C and its annotation were obtained from The Saccharomyces Genome Database (SGD). Quantitative data on protein abundance levels in Saccharomyces cerevisiae S288C cells were taken from the Protein Abundance Database (PaxDb). Pre-trained BERT-like transformers GENA-LM and ESM-2 were used to obtain embeddings of gene and protein sequences, respectively. These embeddings are used as input vectors for training a gradient boosting model to predict protein abundance.

Funding: The study was funded by the Ministry of Science and Higher Education of the Russian Federation project “Kurchatov Center for World-Class Genomic Research” No. 075-15-2019-1662 from 2019-10-31.

doi 10.18699/bgrs2024-8.4-03

Сравнение систем защиты свободноживущих прокариот, экстремофилов и внутриклеточных паразитов

Ириоглов Р.[1, 2]*, Алексеевский А.[1, 2, 3]

[1] *НИИ Физико-химической биологии имени А.Н. Белозерского МГУ, Москва, Россия*

[2] *Факультет биоинженерии и биоинформатики МГУ, Москва, Россия*

[3] *ФГУ ФНЦ Научно-исследовательский институт системных исследований РАН, Москва, Россия*

* *irio-roman@yandex.ru*

Ключевые слова: внутриклеточные паразиты; системы защиты; бактериофаги; бактерии

Мотивация и цель: Бактерии и фаги находятся в постоянном противостоянии: первые развивают и совершенствуют системы защиты, вторые – разрабатывают методы их обхода. Изучение бактериофагов и систем защиты бактерий расширило понимание природы вирусной инфекции, стало основой для установления молекулярных механизмов наследственности, а также дало и продолжает давать генной инженерии новые эффективные инструменты и методики. Целью нашей работы было сравнить состав защитных систем у преимущественно свободноживущих бактерий и внутриклеточных паразитов, а также проверить защитные системы прокариот, обитающих в экстремальных средах.
Материалы: Список видов прокариот и число исследованных геномов штаммов каждого вида представлены в табл. 1.

Таблица 1. Список видов прокариот и число исследованных геномов штаммов каждого вида

Вид	Число штаммов	Вид	Число штаммов
Brucella sp.	150	*Chlamydia trachomatis*	30
Xanthomonas fragariae	30	*Chlamydia psittaci*	30
Stenotrophomonas maltophilia	30	*Shigella flexneri*	30
Geobacter sulfurreducens	30	*Shigella sonnei*	30
Microcystis aeruginosa	8	*Mycobacterium leprae*	5
Mycobacterium tuberculosis	8	*Deinococcus radiodurans*	14
Clostridium botulinum	30	*Thermus thermophiles*	10
Clostridium perfringens	30	*Desulforudis audaxviator*	4
Staphylococcus aureus	30	*Halobacterium salinarum*	6
Rickettsia japonica	30		

Методы и алгоритмы: Для поиска основных компонентов систем защиты в геномах прокариот мы использовали HMM профили, доступные в виде онлайн-сервиса PADLOC [1, 2]. PADLOC способен обнаруживать системы, основанные

на модификации ДНК, абортивные и токсин-антитоксиновые системы, системы CRISPR-Cas и прокариотических белков Argonaut, а также большинство недавно открытых систем с неописанным/предполагаемым механизмом действия.
Результаты: С помощью сервиса PADLOC мы исследовали противофаговые системы бактерий рода *Brucella*. Нам удалось найти всего две защитные системы: систему рестрикции-модификации типа I и дГТФазную систему, встречаются у 100 % штаммов, что оказалось меньше ожидаемого результата, с учетом огромного разнообразия систем защиты и их частой локализации в форме «защитных островов» – участков, способных к горизонтальному переносу.

Таблица 2. Краткая характеристика систем защиты, найденных сервисом PADLOC у бактерий с разными способами существования

Бактерии	Способ существования	Системы защиты на геном			
		Min	Max	Сред.	Самые представленные
Stenotrophomonas maltophilia	Свободноживущий	4	13	10	RM тип II, Zorya тип III, Gabija
Geobacter sulfurreducens	Свободноживущий	6	15	9	RM тип I, dGTPase, CRISPR-Cas
Microcystis aeruginosa	Свободноживущий	26	39	31	RM тип II, PD-T4-6
Xanthomonas fragariae	Есть контакт с внешней средой	7	10	8	RM тип I, Zorya тип III
Micobacterium tuberculosis	Есть контакт с внешней средой	10	14	12	PD-T4-6, dGTPase
Clostridium botulinum	Есть контакт с внешней средой	3	11	9	RM тип I, PD-T4-6, CRISPR-Cas
Clostridium perfringens	Есть контакт с внешней средой	3	12	8	PD-T4-6, Zorya, viperin
Staphylococcus aureus	Паразит	1	7	5	RM тип II, Mokosh тип II
Brucella sp.	Паразит	2	2	2	RM тип I, dGTPase
Shigella flexneri	Паразит	2	6	3	RM тип I, Mokosh тип II
Shigella sonnei	Паразит	4	6	4	RM тип II, Mokosh тип II
Rickettsia japonica	Облигатный паразит	2	2	2	SoFic, dGTPase (встречаются у 100 %)
Chlamydia trachomatis	Облигатный паразит	0	0	0	–
Chlamydia psittaci	Облигатный паразит	0	0	0	–
Micobacterium leprae	Облигатный паразит	3	4	4	PD-T4-6, dGTPase (встречаются у 100 %)
Deinococcus radiodurans	Радиорезистентный	6	10	9	RM тип IV, PD-T4-3, PD-T4-6, dGTPase (встречаются у 100 %)
Thermus thermophiles	Термофил	4	8	5	SoFic, PD-T4-6, dGTPase (встречаются у 100 %)
Desulforudis audaxviator	Криофил	9	13	10	CRISPR-Cas, Wadjet тип 2, Hma (встречаются у 100 %)
Halobacterium salinarum	Галофил	1	5	3	SoFic, AbiE

Мы выдвинули гипотезу, что бактерии рода *Brucella* и прочие внутриклеточные паразиты обладают меньшим количеством систем защиты, чем свободноживущие виды, так как среда обитания данных паразитов свободна от бактериофагов. Для проверки гипотезы мы идентифицировали защитные системы в геномах 4 видов прокариот – облигатных внутриклеточных паразитов. Добавлены следующие категории прокариот: паразиты – 4 вида внутриклеточных паразитов, имеющие период в жизненном цикле, когда находятся вне среды с фагами; имеющие постоянный контакт с внешней средой – 4 вида и свободно живущие – 3 вида.
Для сравнения мы решили также проверить защитные системы прокариот, обитающих в других экстремальных средах: *Deinococcus radiodurans* (радиорезистентная бактерия), *Thermus thermophiles* (гипертермофил), *Desulforudis audaxviator* (криофил), *Halobacterium salinarum* (галофил). Результаты нахождения защитных систем в геномах представлены в табл. 2.
У облигатных внутриклеточных паразитов защитные системы либо не были обнаружены, либо находилось в среднем от 2 до 4 систем на геном. Обнаруженные у них системы SoFic, dGTPase и PD-T4-6 состоят из одного белка и могут выполнять сторонние функции, что объясняет наличие этих систем у облигатных паразитов даже после характерной для них редукции генома. У паразитов в среднем от 2 до 5 защитных систем на геном. Наиболее часто встречающиеся классические системы рестрикции-модификации типов II и Ш, Mokosh тип II и dGTPase. Такое малое разнообразие защитных систем паразитов требует дальнейшего изучения. У большинства родов свободноживущих бактерий среднее число защитных систем составляло от 8 до 11 на геном. У бактерий-экстремофилов, кроме Halobacterium, не наблюдается значительного падения числа систем, но снижено их разнообразие.
Выводы: Полученные нами данные свидетельствуют в пользу выдвинутой нами гипотезы, а также в пользу влияния экстремального места обитания на набор противофаговых защитных систем бактерий. Мы планируем изучить отдельные типы систем защиты, а также рассмотреть распределение систем защиты бактерий внутри вида с различными местами обитания штаммов.
Финансирование: Исследование поддержано грантом РНФ № 21-14-00135.

Comparison of defense systems of free-living prokaryotes, extremophiles and intracellular parasites

Irioglov R.[1, 2]*, Alekseevskiy A.[1, 2, 3]

[1] *Belozersky Research Institute of Physical and Chemical Biology, Lomonosov MSU, Moscow, Russia*

[2] *Faculty of Bioengineering and Bioinformatics, Lomonosov MSU, Moscow, Russia*

[3] *Research Institute for System Analysis, RAS, Moscow, Russia*

* *irio-roman@yandex.ru*

Key words: intracellular parasites; defense systems; bacteriophages; bacteria

Motivation and Aim: Bacteria and phages are in constant confrontation: the former develop and improve protection systems, the latter develop methods to circumvent them. The study of bacteriophages and bacterial defense systems expanded the understanding of the nature of viral infection, became the basis for establishing the molecular mechanisms of heredity, and also gave and continues to give genetic engineering new

effective tools and techniques. The aim of our work was to compare the composition of protective systems in predominantly free-living bacteria and intracellular parasites, as well as to test the defense systems of prokaryotes living in extreme environments.
Materials: See Table 1.

Table 1. Prokaryotic species and strains genomes used

Species	Strains No.	Species	Strains No.
Brucella sp.	150	*Chlamydia trachomatis*	30
Xanthomonas fragariae	30	*Chlamydia psittaci*	30
Stenotrophomonas maltophilia	30	*Shigella flexneri*	30
Geobacter sulfurreducens	30	*Shigella sonnei*	30
Microcystis aeruginosa	8	*Mycobacterium leprae*	5
Mycobacterium tuberculosis	8	*Deinococcus radiodurans*	14
Clostridium botulinum	30	*Thermus thermophiles*	10
Clostridium perfringens	30	*Desulforudis audaxviator*	4
Staphylococcus aureus	30	*Halobacterium salinarum*	6
Rickettsia japonica	30		

Methods and Algorithms: To search for the main components of defense systems in the genomes of prokaryotes, we used HMM profiles available as an online service PADLOC [1, 2]. PADLOC detects defense systems based on DNA modifications, abortive (Abi) and toxin-antitoxin (TA) systems, CRISPR-Cas, prokaryotic Argonaut, and many recently identified within defense islands systems with unknown mechanism of action.
Results: Using the PADLOC service, we investigated the antiphage systems of bacteria of the genus Brucella. We managed to find only 2 defense systems, namely type I restriction modification system and dGTPase system. Both found in 100 % of strains studied. It turned out to be less than we expected to find, taking into account the huge variety of defense systems and their tendency to localize in the form of "protection islands" – areas capable of horizontal transfer.
We hypothesized that bacteria of the genus Brucella and other intracellular parasites have fewer defense systems than free-living species, since the habitat of these parasites is free of bacteriophages.
To test this hypothesis, we identified defense systems in the genomes of 4 species of prokaryotic obligate intracellular parasites, 4 species of parasites, i.e. intracellular parasites encounter the environment at some stages of their life cycle, 4 species having tight contact with the external environment and 3 free living species. In addition, we studied defense systems of prokaryotes living in other extreme environments: *Deinococcus radiodurans* (radioresistant bacterium), *Thermus thermophiles* (hyperthermophile), *Desulforudis audaxviator* (cryophile), *Halobacterium salinarum* (halophile).
The results of finding defense systems in genomes are presented in Table 2.
In obligate intracellular parasites, defense systems were either not detected, or there were not more than four of them per genome. The SoFic, dGTPase and PD-T4-6 systems

found in them consist of a single protein and can perform third-party functions, which explains the presence of these systems in obligate parasites even after their characteristic genome reduction. Parasites have an average of 2 to 5 defense systems per genome. The most common are classic restriction-modification systems type 2 and 3, Mokosh type II and dGTPase (it has other functions besides protective ones). Such a small variety of parasite defense systems deserves further study.

Table 2. The defense systems found by the PADLOC service in bacteria with different ways of existence

Bacteria	Modes of existence	Defense systems per genome			
		Min	Max	Aver	The most common systems
Stenotrophomonas maltophilia	Free-living	4	13	10	RM type II, Zorya type III, Gabija
Geobacter sulfurreducens	Free-living	6	15	9	RM type I, dGTPase, CRISPR-Cas
Microcystis aeruginosa	Free-living	26	39	31	RM type II, PD-T4-6
Xanthomonas fragariae	Contact with the external environment	7	10	8	RM type I, Zorya type III
Micobacterium tuberculosis	Contact with the external environment	10	14	12	PD-T4-6, dGTPase
Clostridium botulinum	Contact with the external environment	3	11	9	RM type I, PD-T4-6, CRISPR-Cas
Clostridium perfringens	Contact with the external environment	3	12	8	PD-T4-6, Zorya, viperin
Staphylococcus aureus	Parasite	1	7	5	RM type II, Mokosh type II
Brucella sp.	Parasite	2	2	2	RM type I, dGTPase (found in 100%)
Shigella flexneri	Parasite	2	6	3	RM type I, Mokosh type II
Shigella sonnei	Parasite	4	6	4	RM type II, Mokosh type II
Rickettsia japonica	Obligate parasite	2	2	2	SoFic, dGTPase found in 100%)
Chlamydia trachomatis	Obligate parasite	0	0	0	–
Chlamydia psittaci	Obligate parasite	0	0	0	–
Micobacterium leprae	Obligate parasite	3	4	4	PD-T4-6, dGTPase (found in 100%)
Deinococcus radiodurans	Radioresistant	6	10	9	RM type IV, PD-T4-3, PD-T4-6, dGTPase (found in 100%)
Thermus thermophiles	Thermophile	4	8	5	SoFic, PD-T4-6, dGTPase (found in 100%)
Desulforudis audaxviator	Cryophile	9	13	10	CRISPR-Cas, Wadjet type 2, Hma (found in 100%)
Halobacterium salinarum	Halophile	1	5	3	SoFic, AbiE

In most genera of free-living bacteria, the average number of defense systems ranged from 8 to 11. In extremophile bacteria, except for Halobacterium, there is no significant decrease in the number of systems, but their diversity is reduced.

Conclusion: The data obtained by us indicate in favor of the hypothesis as well as in favor of the influence of extreme habitat on the set of antiphage defense systems of bacteria. We plan to study a separate type of defense systems, as well as the distribution of bacterial defense systems within a species with different habitats of strains.

Funding: The study is supported by the RSF No. 21-14-00135.

Список литературы/References

1. Payne L.J., Meaden S., Mestre M.R., Palmer C., Toro N., Fineran P.C., Jackson S.A. PADLOC: a web server for the identification of antiviral defence systems in microbial genomes. *Nucleic Acids Res*. 2022;50(W1):W541-W550. doi 10.1093/nar/gkac400
2. Payne L.J., Todeschini T.C., Wu Y., Perry B.J., Ronson C.W., Fineran P.C., Nobrega F.L., Jackson S.A. Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. *Nucleic Acids Res.* 2021;49:10868-10878. doi 10.1093/nar/gkab883

doi 10.18699/bgrs2024-8.4-05

На эволюционные стратегии эффективности элонгации трансляции в бактериях существенно влияет наличие нерибосомных пептидов

Матушкин Ю.Г.[1]*, Клименко А.И.[1, 2], Лашин С.А.[1, 2], Колчанов Н.А.[1, 2], Афонников Д.А.[1]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

* *mat@bionet.nsc.ru*

Ключевые слова: нерибосомные пептиды; синтетазы нерибосомных пептидов; эффективность элонгации трансляции; бактерии; аннотация генома

Мотивация и цель: Нерибосомные пептиды играют важную роль в жизнедеятельности бактерий и имеют экстремально широкую область биологической активности. В частности, они действуют как антибиотики, токсины, поверхностно-активные вещества, сидерофоры, также выполняют ряд других специфических функции. Биосинтез этих молекул происходит не на рибосомах, а за счет специальных ферментов, образующих кластеры в бактериальных геномах. Мы предположили, что наличие путей синтеза нерибосомных пептидов является специфической особенностью метаболизма бактерий, которая может затрагивать и другие метаболические процессы клетки, в том числе и трансляционные.

Методы и алгоритмы: Мы провели биоинформационный анализ кластеров биосинтетических генов нерибосомных пептидов (NRP BGCs), полученных из ANTISMASH-DB [1], используя цельногеномные последовательности бактериальных геномов, доступные в NCBI Genbank. В основе анализа лежит метод прогнозирования эффективности элонгации трансляции генов, реализованный в программе EloE [2]. Скрипты статистического и биоинформационного анализа были разработаны на языке Python с использованием программной библиотеки Biopython.

Результаты: У организмов, геномы которых содержат кластеры биосинтеза нерибосомных пептидов существенно меньшая часть генов регулируется количеством локальных инвертированных повторов, большая часть генов регулируется за счет усредненной энергии шпилек инвертированных повторов и дополнительно за счет кодонного состава [3].

Выводы: В работе впервые показана связь между механизмом регуляции трансляции белок-кодирующих генов в бактериях, который в значительной степени определяется эффективностью элонгации трансляции, и наличием в геномах кластеров генов биосинтеза нерибосомных пептидов. Полученные нами результаты позволяют предположить, что наличие биосинтетических путей нерибосомных пептидов в геномах может оказывать влияние на структуру общего метаболизма бактерий, что выражается и в специфике механизмов рибосомного биосинтеза генов.

Финансирование: Исследование поддержано бюджетным проектом № FWNR-2022-0020.

Evolutionary strategies for translation elongation efficiency in bacteria are significantly influenced by the presence of non-ribosomal peptides

Matushkin Yu.G.[1]*, Klimenko A.I.[1, 2], Lashin S.A.[1, 2], Kolchanov N.A.[1, 2], Afonnikov D.A.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *mat@bionet.nsc.ru*

Key words: non-ribosomal peptides; non-ribosomal peptide synthetases; translational elongation efficiency; bacteria; genome annotation

Motivation and Aim: Non-ribosomal peptides play an important role in bacterial viability and have an extremely wide range of biological activities. In particular, they act as antibiotics, toxins, surfactants, siderophores, and also perform a number of other specific functions. The biosynthesis of these molecules does not occur on ribosomes, but through special enzymes that form clusters in bacterial genomes. We hypothesized that the presence of non-ribosomal peptide synthesis pathways is a specific feature of bacterial metabolism that may affect other metabolic processes of the cell, including translational ones.

Methods and Algorithms: We performed bioinformatics analysis of non-ribosomal pethid biosynthetic gene clusters (NRP BGCs) derived from ANTISMASH-DB [1] using whole-genome sequences of bacterial genomes available at NCBI Genbank. The analysis is based on the method of predicting gene translation elongation efficiency implemented in the EloE program [2]. Statistical and bioinformatics analysis scripts were developed in Python using the Biopython program library.

Results: In organisms whose genomes contain clusters of non-ribosomal peptide biosynthesis, a significantly smaller proportion of genes are regulated by the number of local inverted repeats, while the majority of genes are regulated by the average energy of hairpin inverted repeats and additionally by codon composition [3].

Conclusion: This work shows for the first time the relationship between the mechanism of translation regulation of protein-coding genes in bacteria, which is largely determined by the efficiency of translation elongation, and the presence of non-ribosomal peptide biosynthesis gene clusters in genomes. Our results suggest that the presence of non-ribosomal peptide biosynthetic pathways in genomes may influence the structure of the general metabolism of bacteria, which is also expressed in the specificity of the mechanisms of ribosomal gene biosynthesis.

Funding: The study is supported by project (No. FWNR-2022-0020).

Список литературы/References

1. Blin K., Shaw S., Kautsar S.A., Medema M.H., Weber T. The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res.* 2021;49(D1):D639-D643

2. Sokolov V., Zuraev B., Lashin S., Matushkin Y. Web application for automatic prediction of gene translation elongation efficiency. *J Integr Bioinform.* 2015;12(1):256. doi 10.2390/biecoll-jib-2015-256
3. Клименко А.И., Лашин С.А., Колчанов Н.А., Афонников Д.А., Матушкин Ю.Г. Молекулярные механизмы оптимизации элонгации трансляции существенно различаются у бактерий, имеющих и не имеющих кластеры генов биосинтеза нерибосомных пептидов. *Молекулярная биология*. 2023;57(2):155-165
[Klimenko A.I., Lashin S.A., Kolchanov N.A., Afonnikov D.A., Matushkin Yu.G. Molecular mechanisms to optimize gene translation elongation differ significantly in bacteria with and without nonribosomal peptides. *Molekulyarnaya Biologiya*. 2023;57(2):155-165 (in Russian)]

doi 10.18699/bgrs2024-8.4-06

Поиск распространяющихся микроорганизмов, основанный на анализе представленности *k*-меров в данных метагеномного секвенирования

Панова В.*, Попов Н., Федоров Д., Манолов А.
НИИ системной биологии и медицины Роспотребнадзора, Москва, Россия
** panova_vv@sysbiomed.ru*

Ключевые слова: *k*-меры; счетчик *k*-меров; метагеномы; заболевания; пандемия

Мотивация и цель: После пандемии COVID-19 все больше внимания уделяется новым зоонозным инфекциям. Целью данной работы является разработка подхода к идентификации возможных инфекционных агентов на основе увеличения представленности и распространенности коротких подстрок фиксированной длины (*k*-меров) в данных полногеномного метагеномного секвенирования.
Методы и алгоритмы: Для валидации предложенного метода было проведено симулирование метагеномных прочтений при помощи утилиты InSilicoSeq [1]. Набор и значения представленности видов были получены из анализа секвенирования тотальной РНК выделенной из мазка из носоглотки пациента с COVID-19 (эксперимент SRA SRX8697677, запуск SRR12183113). Для подсчета канонических *k*-меров был использован Jellyfish [2]. Для оценки таксономического состава образца использовали Kraken2 [3].
Для проверки применимости подхода, нами была смоделирована временная серия из 7 шагов, выбран один организм (*Streptococcus sp. S5*) представленность которого линейно возрастала. Выбор организма основан на его представленности, близкой к медианному значению, рост относительно представленности производился с исходного значения до 61, 71 или 91 процентиля. Длина *k*-мера при анализе была выбрана равной 10.
Для подбора оптимального значения длины *k*-меров был использован тот же набор изучаемых видов. В симуляцию также был добавлен шум: относительная численность каждого вида была распределена в соответствии с отрицательным биномиальным законом и изменялась в течение семи временных шагов. Количество прочтений данного соответствующего таксона принято за неотрицательное среднее распределение (mu). Длины результирующих *k*-меров были равны 10, 11, 12, 13, 14. Оценка роста производилась с помощью линейной модели (функция lm в языке программирования R).
Результаты: В валидационном наборе данных процент растущих k-меров (k-меров, число которых статистически значимо растет во временной серии) из целевого организма от всех растущих *k*-меров составляет от 84.48 до 85.57 %. В наборе данных, который использовался для подбора оптимального значения длины *k*-меров, процент растущих *k*-меров из целевого организма от всех растущих *k*-меров уменьшается от 99.97 % при $k = 10$ до 52.19 % при $k = 14$. Процент растущих k-меров из всех k-меров целевого организма увеличивается от 1.74 % при $k = 10$ до 82.67 % при $k = 14$. В открытых базах данных (SRA,

HumanMetagenomeDB) собраны метагеномы из России, на которых планируется запустить наш пайплайн.
Выводы: Подход, основанный на анализе динамики *k*-меров, потенциально позволяет искать возможные агенты (вирусы и бактерии), увеличивающиеся в распространенности. Одной из причин увеличения распространенности может быть появление новых инфекционных агентов. Большая длина *k*-мера означает рост ложноположительных случаев, меньшая – ложноотрицательных.
Финансирование: Работа выполнена на базе НИИ системной биологии и медицины Роспотребнадзора в рамках госзадания «Разработка алгоритмов для выявления новых, уникальных последовательностей ДНК или РНК в метагеномах и их фенотипическая характеристика *in vitro*», номер 12203090069-4.

Search for spreading microorganisms based on the analysis of the representation of *k*-mers in metagenomic sequencing data

Panova V.*, Popov N., Fedorov D., Manolov A.
Research Institute for Systems Biology and Medicine (RISBM), Moscow, Russia
** panova_vv@sysbiomed.ru*

Key words: k-mers; k-mer counter; metagenomes; diseases; pandemic

Motivation and Aim: After the COVID-19 pandemic, more and more attention is being paid to new zoonotic infections. This work aims to develop an approach to the identification of possible infectious agents based on increasing the representation and prevalence of short fixed-length substrings (*k*-mers) in the data of genome-wide metagenomic sequencing.
Methods and Algorithms: To validate the proposed method, metagenomic reads were simulated using the InSilicoSeq tool [1]. The set and values of species representation were obtained from the analysis of sequencing of total RNA isolated from a nasopharyngeal swab of a patient with COVID-19 (experiment SRA SRX8697677, launch SRR12183113). Jellyfish was used to calculate canonical *k*-mers [2]. Kraken2 was used to assess the taxonomic composition of the sample [3].
To test the applicability of the approach, we modeled a time series of 7 steps and selected one organism (*Streptococcus sp. S5*) whose abundance increased linearly. The choice of an organism is based on its abundance close to the median value; an increase in relative representation was made from the initial value to 61, 71, or 91 percentiles. The length of the *k*-mer was chosen to be 10 in the analysis.
The same set of studied species was used to select the optimal value of the length of *k*-mers. The noise was also added to the simulation: the relative abundance of each species was distributed according to the negative binomial law and varied over seven-time steps. The number of readings of this corresponding taxon is taken as a non-negative mean distribution (mu). The lengths of the resulting *k*-mers were equal to 10, 11, 12, 13, 14. The growth was estimated using a linear model (the lm function in the R programming language).
Results: In the validation dataset, the percentage of growing *k*-mers (*k*-mers, the number of which grows statistically significantly in the time series) from the target organism of all growing *k*-mers ranges from 84.48 to 85.57 %. In the dataset that was used to select the optimal value for the length of *k*-mers, the percentage of growing *k*-mers from the

target organism of all growing k-mers decreases from 99.97 % at $k = 10$ to 52.19 % at $k = 14$. The percentage of growing k-mers from all k-mers of the target organism increases from 1.74 % at $k = 10$ to 82.67 % at $k = 14$.
Metagenomes from Russia were collected from open databases (SRA, HumanMetagenomeDB). Our pipeline is planned to be launched on these metagenomes.
Conclusion: The approach based on the analysis of the dynamics of *k*-mers potentially allows us to search for possible agents (viruses and bacteria) that are increasing in prevalence. A longer *k*-mer length means an increase in false-positive cases, and a smaller one – false-negative ones.
Funding: The research was carried out on the basis of the Research Institute of Systems Biology and Medicine within the framework of the state task "Development of algorithms for identifying new, unique DNA or RNA sequences in metagenomes and their phenotypic characteristics *in vitro*", number 12203090069-4.

Список литературы/References

1. Gourlé H., Karlsson-Lindsjö O., Hayer J., Bongcam-Rudloff E. Simulating Illumina metagenomic data with InSilicoSeq. *Bioinformatics.* 2019;35(3):521-522. doi 10.1093/bioinformatics/bty630
2. Marçais G., Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27(6):764-770. doi 10.1093/bioinformatics/btr011
3. Wood D.E., Lu J., Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019;20(1):257. doi 10.1186/s13059-019-1891-0

doi 10.18699/bgrs2024-8.4-07

Поиск и описание разнообразия, эволюции систем рестрикции–модификации в гипертермофильных бактериях и археях

Попова А.[1]*, Русинов И.[1], Алексеевский А.[1, 2]

[1] *Отдел математических методов в биологии НИИ ФХБ им. Белозерского МГУ, Москва, Россия*

[2] *НИИСИ РАН, Москва, Россия*

* *diatrimma@mail.ru*

Ключевые слова: Системы рестрикции-модификации; археи; бактерии; термозимы; эволюция

Мотивация и цель: Гипертермофильные бактерии и археи интересны для науки не только, как часть биоразнообразия живой природы. Ферменты, синтезируемые гипертермофилами, также называемые термозимами, термостабильны и максимально активны при высоких температурах 50–80 °C. Термозимы катализируют те же реакции, что и их мезофильные аналоги. Это позволяет применять их в различных экспериментах и лабораторных методиках и промышленных процессах, требующих поддержания высоких температур, которые не выдерживают мезофильные ферменты.

Эндонуклеазы рестрикции ферменты важные для молекулярной биологии, и термостабильные не являются исключением. Например, в работе Bonanno et al. 2007 года [1] предложен метод FLAG для выявления метилирования конкретных сайтов CpG в промоторах онкогенов. FLAG (генерация флуоресцентных ампликонов) представляет собой технологию генерации гомогенного сигнала, основанную на исключительно термостабильной эндонуклеазе PspGI. FLAG обеспечивает генерацию сигнала в реальном времени во время ПЦР путем PspGI-опосредованного расщепления погашенных флуорофоров на 5'-конце двухцепочечных продуктов ПЦР [1].

Помимо возможного практического применения термостабильные системы рестрикции-модификации (Р-М) интересны своей эволюцией, составом и распространенностью в контексте защитных систем прокариот.

Цель работы заключалась в изучении эволюции, родства, состава и распространенности систем Р-М в 7 родах гипертермофилов: *Thermocrinis*, *Sulfolobus*, *Pyrobaculum*, *Hyperthermus*, *Thermus*, *Thermotoga*, *Thermococcus*.

Методы и алгоритмы: Данные о системах рестрикции-модификации были получены из открытой базы данных REBASE (обновление от 6 июня 2023 г.). Белковые последовательности систем рестрикции-модификации выбранных нами семи родов гипертермофилов обрабатывались по отдельности в зависимости от типа белка: R\N – эндонуклеазы рестрикции и никазы, S – субъединицы S, узнающие последовательность, M – метилтрасферазы с помощью программы MEGA 11.

Последовательности анализировались с помощью построения множественных выравниваний и филогенетических деревьев. Были построены таблицы принадлежности систем штаммам и распространенности различных типов систем среди каждого вида принадлежащих 7 родам отобранных нами гипертермофилов.

Результаты: Среди систем были распространены все 4 типа.
Тип I. Эти ферменты представляют собой многосубъединичные белки, которые функционируют как единый белковый комплекс и обычно имеют типичный состав генов M|S|R (M – Мтаза, S – субъединица, узнающая сайт ДНК, R – ЭР). После узнавания мишени ферментативный комплекс перемещается по молекуле ДНК до момента встречи с каким-либо препятствием. Таким образом, разрыв вносится на большом произвольном расстоянии от сайта рестрикции. Для расщепления ДНК СР-М типа I необходим гидролиз АТФ. Это связано с тем, что для перемещения по молекуле ДНК и ее расплетания после распознавания сайта СР-М требуется энергия.
Тип II. Системы рестрикции типа II обычно имеют состав: R|M, R|M|M, RM (подтип IIG, один белок включает ЭР и МТаз каталитические домены). Эндонуклеазы разрезают ДНК в строго определенных позициях внутри или около сайта рестрикции. Метилтрансферазы обычно функционируют в форме мономера.
Р-М Типа III включают две субъединицы (Res и Mod), которые объединяются в гетеротетрамер (Res2Mod2) с эндонуклеазной и метилтрансферазной активностями. Системы типа III имеют состав R|M, отличие от типа II в необходимости двух близко расположенных сайтов для расщепления ДНК ЭР и механизме действия ЭР.
Тип IV. Система Тип IV имеет состав R или R|R, состоящие из одного или двух генов, кодирующих белки, которые расщепляют только модифицированную ДНК, включая метилированные, гидроксиметилированные и глюкозил-гидроксиметилированные
Одиночные МТазы (в REBASE отнесены к системам типа II с составом M) также распространена во всех видах [2].
Кроме перечисленных систем, встречались системы R-M с нетипичными составами M|S|S|R – 6 систем, M|S|S|S|R – 1, M|V – 5 (V – никаза, расщепляющая одну цепочку ДНК), M|R|V – 3, M|S – 5, M|M – 3, S – 3, S|R – 1. В дальнейшей работе планируется разобраться, являются ли эти случаи ошибками предсказания или же это особые системы.
Была составлена таблица, содержащая 65 видов, 174 штаммов и 578 систем. Построены сводные таблицы представленности по видам и родам. В дальнейшем, планируется посмотреть представленность среди гипертермофилов систем Р-М согласно классификации Р-М по каталитическим доменам ЭР и Мтаз, которая разработана в текущих лабораторных исследованиях.
Выводы: В роде *Thermococcus* самыми распространенными являлись системы первого типа, в роде *Thermotoga* самыми распространенными были системы состава R|M, в роде *Thermus* самыми распространенными были системы состава R|M и RM, *Thermocrinis*, *Sulfolobus*, *Pyrobaculum* самыми распространенными были системы состава R|M.
Нетипичные системы встречались чаще всего у штаммов вида *Thermus thermophilus* и штаммов с неопределенным видом у рода *Thermococcus*.
Финансирование: Исследование поддержано грантом РНФ № 21-14-00135.

Search and description of diversity, evolution of restriction–modification systems in hyperthermophilic bacteria and archaea

Popova A.[1]*, Rusinov I.[1], Alexeevski A.[1, 2]

[1] *Department of Mathematical Methods in Biology A.N. Belozersky Research Institute of Physico-Chemical Biology MSU, Moscow, Russia*
[2] *FSI FSC SRISA RAS, Moscow, Russia*
** diatrimma@mail.ru*

Key words: Restriction modification systems; archaea; bacteria; thermowinters; evolution

Motivation and Aim: Hyperthermophilic bacteria and archaea are of interest to science not only as part of the biodiversity of living nature. Enzymes synthesized by hyperthermophiles, also called thermozymes, are thermostable and are most active at high temperatures of 50–80 °C. Thermozymes catalyze the same reactions as their mesophilic counterparts. This allows them to be used in a variety of experiments and laboratory techniques and industrial processes that require maintaining high temperatures that mesophilic enzymes cannot withstand.
Restriction endonucleases are important enzymes in molecular biology, and heat-stable enzymes are no exception. For example, Bonanno et al. In 2007, the FLAG method was proposed to detect methylation of specific CpG sites in oncogene promoters. FLAG (Fluorescent Amplicon Generation) is a homogeneous signal generation technology based on the exclusively thermostable endonuclease PspGI. FLAG enables real-time signal generation during PCR by PspGI-mediated cleavage of quenched fluorophores at the 5' end of double-stranded PCR products [1].
In addition to their possible practical applications, thermostable restriction-modification (RM) systems are interesting for their evolution, composition, and prevalence in the context of prokaryotic defense systems.
The purpose of the work was to study the evolution, relationship, composition and prevalence of R-M systems in 7 genera of hyperthermophiles: *Thermocrinis*, *Sulfolobus*, *Pyrobaculum*, *Hyperthermus*, *Thermus*, *Thermotoga*, *Thermococcus*.
Methods and Algorithms: Data on restriction modification systems were obtained from the open database REBASE (updated June 6, 2023). Protein sequences of restriction-modification systems of the seven genera of hyperthermophiles selected by us were processed separately depending on the type of protein: R\N – restriction endonucleases and nickases, S – subunits S that recognize the sequence, M – methyltransferases using the MEGA 11 program.
Sequences were analyzed using multiple alignments and phylogenetic trees. Tables were constructed of the affiliation of systems to strains and the prevalence of various types of systems among each species belonging to the 7 genera of hyperthermophiles we selected.
Results: All 4 types were common among systems.
Type I. These enzymes are multi-subunit proteins that function as a single protein complex and usually have a typical M|S|R gene composition (M – Mtase, S – DNA site recognition subunit, R – ER). After recognizing the target, the enzymatic complex moves along the DNA molecule until it encounters some obstacle. Thus, the break is introduced at a large random distance from the restriction site. DNA cleavage by CP-M type I

requires ATP hydrolysis. This is due to the fact that energy is required to move along the DNA molecule and unwind it after recognition of the CP-M site.
Type II. Type II restriction systems typically have the following composition: R|M, R|M|M, RM (sub-type IIG, one protein includes ER and MTase catalytic domains). Endonucleases cut DNA at strictly defined positions within or near the restriction site. Methyltransferases typically function in monomer form.
Type III PMs include two subunits (Res and Mod), which combine into a heterotetramer (Res2Mod2) with endonuclease and methyltransferase activities. Type III systems have an R|M composition, differing from type II in the need for two closely located sites for ER DNA cleavage and the mechanism of action of ER.
Type IV. The Type IV system has an R or R|R composition, consisting of one or two genes encoding proteins that cleave only modified DNA, including methylated, hydroxymethylated and glucosyl-hydroxymethylated
Single MTases (referred to as type II systems with M composition in REBASE) are also common in all species [2].
In addition to the listed systems, there were systems of the R-M system with atypical compositions: M|S|S|R – 6 systems, M|S|S|S|R – 1, M|V – 5 (V – nickase, cleaving one DNA chain), M|R|V – 3, M|S – 5, M|M – 3, S – 3, S|R – 1. Future work is planned to determine whether these cases are prediction errors or whether they are special systems.
A table was compiled containing 65 species, 174 strains and 578 systems. Summary tables of representation by species and genera were constructed. In the future, it is planned to look at the representation of R-M systems among hyperthermophiles according to the R-M classification according to the catalytic domains ER and Mt3, which was developed in ongoing laboratory studies.
Conclusion: In the genus *Thermococcus* the most common systems were the first type, in the genus *Thermotoga* the most common systems were the composition R|M, in the genus *Thermus* the most common systems were the composition R|M and RM, *Thermocrinis*, *Sulfolobus*, Pyrobaculum the most common systems were the composition R| M.
Atypical systems were most often found in strains of the species *Thermus* thermophilus and strains of uncertain species in the genus *Thermococcus*.
Funding: The study is supported by RSF grant No. 21-14-00135.

Список литературы/References

1. Bonanno C., Shehi E., Adlerstein D., Makrigiorgos G.M. MS-FLAG, a novel real-time signal generation method for methylation-specific PCR. *Clin Chem*. 2007;53(12):2119-2127. doi 10.1373/clinchem.2007.094011
2. Официальный сайт базы данных REBASE [Электронный ресурс]. URL: http://rebase.neb.com/rebase/rebhelp.html (дата обращения: 20.02.2024)
3. Официальный сайт MEGA 11 [Электронный ресурс]. URL: https://www.megasoftware.net/downloads/dload_win_gui

doi 10.18699/bgrs2024-8.4-08

Оптимизация биоинформатической обработки данных 16S секвенирования микробиоты кишечника

Скоповец Е.Я.*, Вертелко В.Р.
РНПЦ детской онкологии, гематологии и иммунологии, Минск, Беларусь
** Skopovets@yandex.ru*

Ключевые слова: биоинформатика; 16S; ДНК; микробиота

Мотивация и цель: В последние десятилетия технологии секвенирования нового поколения существенно повлияли на исследования микробиома человека, позволяя лучше понять и охарактеризовать взаимодействия микробиома и хозяина. Поскольку большинство видов кишечных микробов трудно культивировать, для изучения кишечного микробиома широко применяются технологии секвенирования «нового» поколения, включая 16S рРНК, 18S рРНК, секвенирование внутреннего транскрибируемого спейсера (ITS), метагеномное секвенирование методом дробовика, метатранскриптомное секвенирование и виромное секвенирование. Целью исследования являлась оптимизация анализа данных 16S секвенирования микробиоты кишечника пациентов с онкогематологическими заболеваниями с помощью разработанного программного ресурса MicrobiotaAnalysisToolkit.
Методы и алгоритмы: Материалом для исследования послужили fastq файлы, полученные в результате парноконцевого NGS секвенирования микробиоты кишечника 24 здоровых детей и 43 пациентов с врожденными дефектами иммунной системы (ПИД) с помощью платформы Illumina Miseq (v3 600 циклов). Биоинформатическая обработка данных проводилась с использованием языка Python. В основе работы программного ресурса MicrobiotaAnalysisToolkit – конвейер DADA2. Назначение таксономии проводилось с помощью базы данных SILVA и Greengenes.
Результаты: Метод 16S секвенирования кишечной микробиоты включает множество этапов: забор материала у пациентов, выделение ДНК и подготовку библиотеки для секвенирования, непосредственно секвенирование, биоинформатическую и статистическую обработку [1].
Биоинформационная обработки данных метагеномного секвенирования включает следующие этапы (рис. 1):
1. Удаление адаптеров, фильтрация и демультиплексирование прочтений.
2. Формирование обучающей выборки и последующее обучение модели на основе алгоритма наивного Байеса для классификации таксономической принадлежности
3. Присвоение таксономии с помощью баз данных GreenGenes и SILVA.
4. Построение сводных графиков и таблиц на основе полученных результатов.

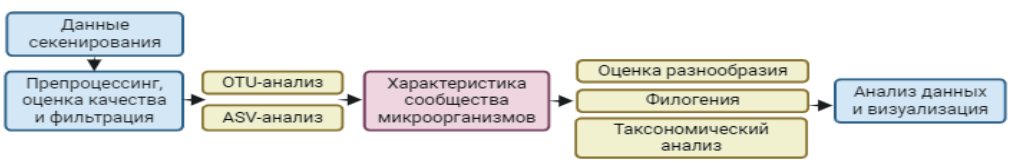


Рис. 1. Алгоритм анализа данных секвенирования 16S рРНК

Для осуществления этапов оценки качества и фильтрации данных секвенирования использован алгоритм DADA2, в частности плагин q2-dada2. Для работы с парноконцевыми данными секвенирования, в плагине q2-dada2 использован подмодуль denoise_paired. При данных вводных параметрах выполняли фильтрацию по качественному признаку, проверяли библиотеку на наличие химерных последовательностей и объединяли парноконцевые прочтения в единый пул данных.

Для работы подмодуля denoise_paired требовалась настройка параметров, которые подобрали эмпирически. Для используемого набора данных (более чем из 30 пациентов) удалось выделить ряд параметров, которые позволили улучшить качественную характеристику набора данных: p-trim-left-f и p-trim-left-r (параметры указывающие сколько нуклеотидов необходимо убрать с начала каждого фрагмента, для прямых и обратных прочтений соответственно), p-trunc-len-f и p-trunc-len-r (параметры указывающие сколько нуклеотидов необходимо оставить, для прямых и обратных прочтений соответственно), p-min-fold-parent-over-abundance (параметр указывающий минимальное количество перекрывающихся нуклеотидов, используется при объединении парноконцевых прочтений), p-max-ee-f и p-max-ee-r (параметры определяющие максимальное количество возможных несоответствий при объединении прочтений). В дальнейшем использовали обученную модель на основе алгоритма наивного Байеса для классификации таксономической принадлежности сформированных ранее последовательностей.

Завершающий этап присвоение таксономии проводили с использованием двух доступных баз данных: SILVA или Greengenes. Определение эталонной базы данных может существенно повлиять на все результаты анализа: процесс группировки/кластеризации бактерий в значительной степени привязан к выбранной эталонной базе данных, что напрямую влечет за собой риск получения противоречивых результатов при прямом сравнении (рис. 2) [2, 3].

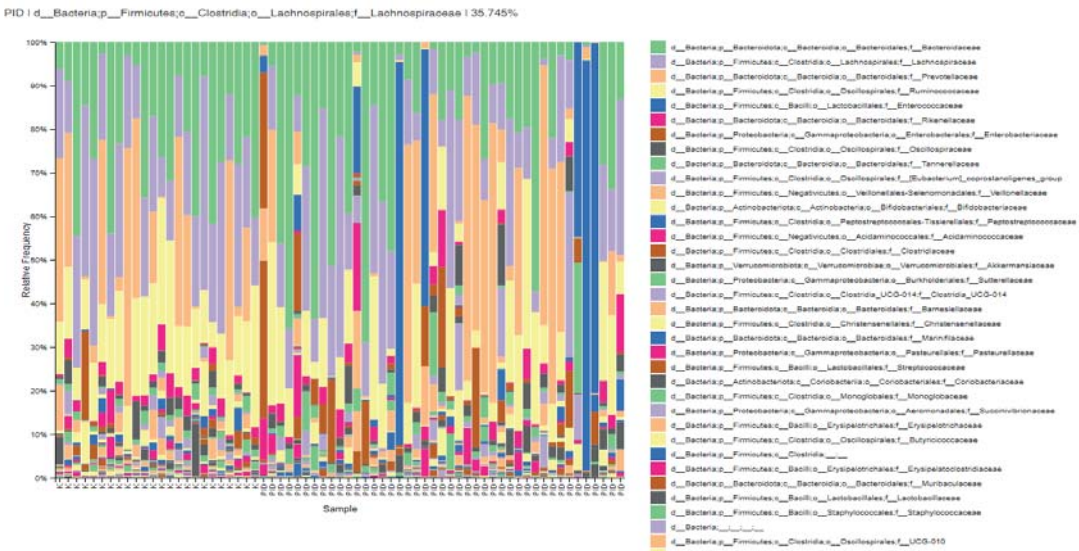


Рис. 2. Таксономический состав микробиоты кишечника на уровне семейства здоровых лиц и пациентов с врожденными дефектами иммунной системы

Процент идентичности группировки также играет важную роль, поскольку и Greengenes, и SILVA обеспечивают разные уровни идентичности, при этом 97 % является наиболее популярным и общепринятым: считается, что штаммы, принадлежащие к одному и тому же виду бактерий, имеют разницу < 3 % в их последовательности 16S рРНК.
Выбор базы данных влияет не только на таксономический состав (особенно на уровне рода), но и на все филогенетические оценки, как на метрики альфа-разнообразия, так и на расстояния бета-разнообразия, поскольку каждая база данных имеет свой определенный множественно-выровненный эталонный набор в качестве руководства по составлению филогенетического дерева.
Следующий шаг – это вычисление ряда общих показателей разнообразия в нашей функциональной таблице. Для этого использован плагин core-metrics-phylogenetic и пайплайн, представляющий собой набор различных команд для QIIME2.
Выводы: Для осуществления процедуры метагеномного анализа, необходимо обладать знаниями в области медицины и программирования, системного администрирования, а также ряд узкоспециализированных навыков из области биоинформатики, математического моделирования и статистики, что в значительной степени увеличивает необходимый уровень квалификации специалиста. В свою очередь, использование предлагаемого веб-ресурса, позволяет автоматизировать все описанные ранее этапы обработки – в значительной степени снизив уровень вхождения и, что немаловажно, повысив удобство и скорость обработки, так как конечному пользователю необходимо лишь загрузить данные через графический интерфейс, выбрать из встроенного перечня необходимые параметры и нажать кнопку запуска.

Optimization of bioinformatics analysis of 16S sequencing data of gut microbiota

Skopovets K.Ya.*, Vertelko V.R.
Belarusian Research Center for Pediatric Oncology, Hematology and Immunology, Minsk, Belarus
** Skopovets@yandex.ru*

Key words: bioinformatics; 16S; DNA; microbiota

Motivation and Aim: In recent decades, next-generation sequencing technologies have significantly impacted human microbiome research, allowing for a better understanding and characterization of microbiome-host interactions. Because most gut microbial species are difficult to culture, “next-generation” sequencing technologies, including 16S rRNA, 18S rRNA, internal transcribed spacer (ITS) sequencing, metagenomic shotgun sequencing, metatranscriptome sequencing, and virome sequencing, have been widely used to study the gut microbiome. The purpose of the study was to optimize the analysis of 16S sequencing data of the intestinal microbiota of patients with oncohematological diseases using the developed software resource MicrobiotaAnalysisToolkit.
Methods and Algorithms: The material for the study was fastq files obtained as a result of paired-end NGS sequencing of the intestinal microbiota of 24 healthy children and 43 patients with inborn errors of immunity (PID) using the Illumina Miseq platform (v3 600 cycles). Bioinformatics data processing was carried out using the Python language.

The operation of the MicrobiotaAnalysisToolkit software resource is based on the DADA2 pipeline. Taxonomy assignment was done using the SILVA and Greengenes database.

Results: The 16S intestinal microbiota sequencing method includes many stages: collecting material from patients, DNA extraction and library preparation for sequencing, sequencing itself, bioinformatics and statistical processing [1].

Bioinformatics processing of metagenomic sequencing data includes the following steps (Fig. 1):

1. Removing adapters, filtering and demultiplexing reads
2. Formation of a training sample and subsequent training of the model based on the Naive Bayes algorithm for taxonomic classification
3. Taxonomy assignment using GreenGenes and SILVA databases.
4. Construction of summary graphs and tables based on the results obtained.

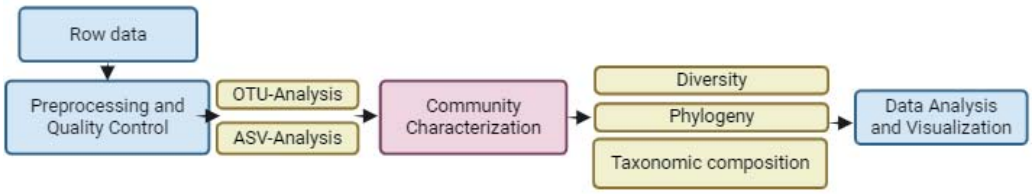


Fig. 1. Algorithm for analysis 16S rRNA sequencing data

To carry out the stages of quality assessment and filtering of sequencing data, the DADA2 algorithm, in particular the q2-dada2 plugin, was used. To work with paired-end sequencing data, the q2-dada2 plugin uses the denoise_paired submodule. With these input parameters, filtering was performed based on qualitative criteria, the library was checked for the presence of chimeric sequences, and paired-end reads were combined into a single data pool.

For the denoise paired submodule to work, it was necessary to configure parameters that were selected empirically. For the data set used (more than 30 patients), it was possible to identify a number of parameters that made it possible to improve the qualitative characteristics of the data set: p-trim-left-f and p-trim-left-r (parameters indicating how many nucleotides need to be removed from the beginning of each fragment, for forward and reverse reads, respectively), p-trunc-len-f and p-trunc-len-r (parameters indicating how many nucleotides need to be left, for forward and reverse reads, respectively), p-min-fold-parent-over- abundance (a parameter indicating the minimum number of overlapping nucleotides, used when merging paired-end reads), p-max-ee-f and p-max-ee-r (parameters defining the maximum number of possible mismatches when merging reads). Subsequently, a trained model based on the Naive Bayes algorithm was used to classify the taxonomic affiliation of the previously generated sequences.

The final stage of taxonomy assignment was carried out using two available databases: SILVA or Greengenes. The definition of a reference database can have a significant impact on the overall results of the analysis: the process of grouping/clustering bacteria is largely tied to the selected reference database, which directly entails the risk of producing inconsistent results in direct comparisons (Fig. 2) [2, 3].

Grouping percentage identity also plays an important role, as both Greengenes and SILVA provide different levels of identity, with 97 % being the most popular and generally accepted: strains belonging to the same bacterial species are considered to have < 3 % difference in their 16S rRNA sequences.

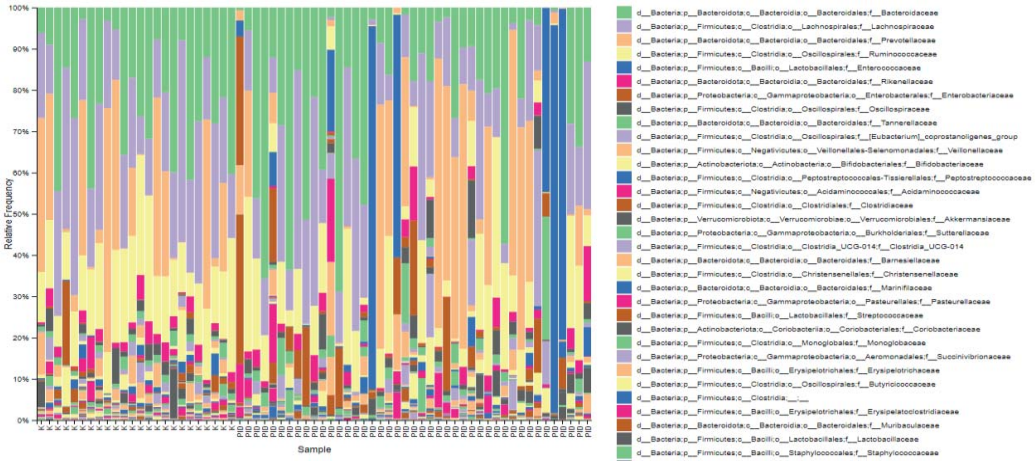


Fig. 2. Taxonomic composition of the intestinal microbiota at the family level of healthy individuals and patients with inborn errors of immunity

The next step is to calculate a number of common diversity indicators in our functional table. For this, the core-metrics-phylogenetic plugin and pipeline, which is a set of various commands for QIIME2, were used.

Conclusion: To carry out the metagenomic analysis procedure, it is necessary to have knowledge in the field of medicine and programming, system administration, as well as a number of highly specialized skills in the field of bioinformatics, mathematical modeling and statistics, which significantly increases the required level of specialist qualifications. In turn, the use of the proposed web resource allows you to automate all the previously described processing stages – significantly reducing the level of entry and, importantly, increasing the convenience and speed of processing, since the end user only needs to download data through a graphical interface, select from the built-in list the required parameters and press the start button.

Список литературы/References

1. Gao B., Chi L., Zhu Y., Shi X., Tu P., Li B., Yin J., Gao N., Shen W., Schnabl B. An introduction to next generation sequencing bioinformatic analysis in gut microbiome studies. *Biomolecules*. 2021;11(4):530. doi 10.3390/biom11040530
2. Yen S., Johnson J.S. Metagenomics: a path to understanding the gut microbiome. *Mamm Genome*. 2021;32(4):282-296. doi 10.1007/s00335-021-09889-x
3. Liu Y.X., Qin Y., Chen T., Lu M., Qian X., Guo X., Bai Y. A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell*. 2021;12(5):315-330. doi 10.1007/s13238-020-00724-8

doi 10.18699/bgrs2024-8.4-09

Studying of the effect of THz radiation on cells of the thermophilic bacterium *Geobacillus icigianus*

Bannikova S.V.[1, 2]*, Khlebodarova T.M.[1, 2], Oshchepkov D.Y.[1, 2], Shipova A.A.[1, 2], Uvarova Y.E.[1, 2], Vasilieva A.R.[1, 2], Vasiliev G.V.[1, 2], Mukhin A.M.[1, 2], Bryanskaya A.V.[1, 2], Popik V.M.[3], Peltek S.E.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Budker Institute of Nuclear Physics, SB RAS, Novosibirsk, Russia*

* *sbann@bionet.nsc.ru*

Key words: terahertz radiation; *Geobacillus icigianus*; transcriptome

Motivation and Aim: As the applications of non-thermal terahertz (THz) radiation expand, especially in biomedicine, evidence is accumulating that this type of radiation can influence both biological molecules directly and cellular processes in general. Moreover, there is evidence of changes in intracellular processes under the influence of THz radiation at the transcription level. Thus, in the *E. coli* cell, THz irradiation changes the expression of genes involved in protection from oxidative stress, maintaining the homeostasis of transition metals, amino acid metabolism, etc. [1–4]. However, identifying the biological effects of THz radiation in thermosensitive organisms is associated with a number of technical difficulties, one of which is maintaining optimal temperature [5]. This study analyzed the effect of THz radiation on gene expression in the cells of the thermophilic bacterium *Geobacillus icigianus*, which is adapted to a wide temperature range from 50 to 75 °C [6]. It was previously noted that THz irradiation did not lead to the appearance of heat shock proteins in its cell [7].

Methods and Algorithms: The experiments were carried out on the basis of a terahertz free electron laser at the Siberian Center for Synchrotron and Terahertz Radiation (Institute of Nuclear Physics SB RAS). This research was supported in part through computational resources of HPC facilities at collaborative center “Bioinformatics” ICG SB RAS. Cells in LB medium in a specially designed cuvette were irradiated for 15 min with THz radiation with a power density of 0.23 W/cm^2 and a wavelength of 130 μm. To identify genes that change their expression level under the influence of THz radiation, the results of previously conducted whole-genome sequencing of the *G. icigianus* genome were used. Analysis of the effect of THz radiation on the cell transcriptome was carried out using RNA sequencing.

Results: A comparative analysis of the transcriptome of *G. icigianus* cells immediately after THz irradiation (2.3 THz, 15 min.) and 10 min after the end of irradiation showed the absence of activation of heat shock protein genes and showed a negative effect of THz irradiation on the gene systems that control the growth of the bacterium and the structure of its cellular wall, demonstrated inhibition of transcription of a number of genes, the products of which are involved in redox reactions and protection from oxidative stress, genes of the chaperone protein ClpB and the DNA repair system protein RadA, as well as genes of the two-component system for phosphorylation of arginine residues using McsB kinase. The latter is critical in regulating the cellular response to oxidative stress, marking proteins and their subsequent degradation by the Clp protease,

as well as regulating processes such as cell motility and competence. At the same time, activation of transcription of genes for the metabolism of sugars, fructose and xylose, synthesis of lactic acid, transport of oligosaccharides and cyclodextrins was observed, which may be a cell response to THz irradiation associated with a general inhibition of metabolism observed during this period at the proteomic level [7], and aimed at using additional carbon sources. The most sensitive to THz irradiation were the gene systems for maintaining the homeostasis of transition metals – copper, iron and zinc, the inhibition of the expression of which increased significantly 10 min after the end of exposure. Blocking the expression of genes for the copper export and iron import systems under the influence of THz irradiation may be a precursor to the development of toxic stress in *G. icigianus* cells, leading to destabilization of the function of proteins containing Fe/S clusters in their structure.

Conclusion: Transcriptomic analysis of irradiated *G. icigianus* cells confirmed the non-thermal nature of THz radiation [7], as well as the sensitivity of gene systems for maintaining transition metal homeostasis to irradiation, also observed in *E. coli* [2]. Unlike *E. coli* [1], cells of the thermophilic bacterium *G. icigianus* demonstrated rapid restoration of the activity of gene systems for anti-stress protection of DNA structure and proteins after THz irradiation.

Funding: The study is supported from the Kurchatov Genomic Center of the Institute of Cytology and Genetics SB RAS (No. 075-15-2019-1662) and the Ministry of Science and Higher Education of the Russian Federation (No. FWNR-2022-0022).

References

1. Demidova E.V., Goryachkovskaya T.N., Malup T.K., Bannikova S.V., Semenov A.I., Vinokurov N.A., Kolchanov N.A., Popik V.M., Peltek S.E. Studying the non-thermal effects of terahertz radiation on E. coli/pKatG-GFP biosensor cells. *Bioelectromagnetics*. 2013;34(1):15-21. doi 10.1002/bem.21736
2. Demidova E.V., Goryachkovskaya T.N., Mescheryakova I.A., Malup T.K., Semenov A.I., Vinokurov N.A., Kolchanov N.A., Popik V.M., Peltek S.E. Impact of terahertz radiation on stress-sensitive genes of E. coli cell. *IEEE Transact Terahertz Sci Technol*. 2016;6(3):435-441
3. Serdyukov D.S., Goryachkovskaya T.N. et al. Study on the effects of terahertz radiation on gene networks of Escherichia coli by means of fluorescent biosensors. *Biomed Opt Express*. 2020;11(9):5258-5273. doi 10.1364/BOE.400432
4. Serdyukov D.S., Goryachkovskaya T.N., Mescheryakova I.A., Kuznetsov S.A., Popik V.M., Peltek S.E. Fluorescent bacterial biosensor E. coli/pTdcR-TurboYFP sensitive to terahertz radiation. *Biomed Opt Express*. 2021;12(2):705-721. doi 10.1364/BOE.412074
5. Sun L., Zhao L., Peng R.Y. Research progress in the effects of terahertz waves on biomacromolecules. *Mil Med Res*. 2021;8(1):28. doi 10.1186/s40779-021-00321-8
6. Bryanskaya A.V., Rozanov A.S., Slynko N.M., Shekhovtsov S.V., Peltek S.E. Geobacillus icigianus sp. nov., a thermophilic bacterium isolated from a hot spring. *Int J Syst Evol Microbiol*. 2015;65(Pt. 3):864-869. doi 10.1099/ijs.0.000029
7. Bannikova S., Khlebodarova T., Vasilieva A., Mescheryakova I., Bryanskaya A., Shedko E., Popik V., Goryachkovskaya T., Peltek S. Specific features of the proteomic response of thermophilic bacterium Geobacillus icigianus to terahertz irradiation. *Int J Mol Sci*. 2022;23(23):15216. doi 10.3390/ijms232315216

doi 10.18699/bgrs2024-8.4-10

Assembly and annotation pipeline for microorganism genomes from sequence to gene networks

Demenkov P.S.[1, 2]*, Lashin S.A.[1, 2], Ivanisenko V.A.[1, 2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *demps@bionet.nsc.ru*

Microorganisms of natural and technogenic ecosystems represent an inexhaustible pool of metabolic pathways for the utilization of some compounds and the biosynthesis of others. Modern experimental technologies of genetics and microbiology provide an effective search in microbiological collections for new strains of microorganisms that are promising for biotechnological applications, as well as complete sequencing of their genomes. The practice of the world's major biotechnology companies, accumulated over the past 30 years, indicates the critical importance of integrating experimental and information-computer approaches for constructing strains of microorganisms with target properties, including multi-parameter profiling of microorganisms based on genomic, proteomic and other data, bioinformatics and systems methods computer biology, which allows, on the basis of genomic information, to reconstruct gene networks that control the production of target substances and to build mathematical models of their functioning.

This paper proposes a software pipeline for assembly, annotation of sequenced genomes of microorganisms and reconstruction of their gene networks. The pipeline includes the following steps: 1) Quality control and sequence cleaning (FastQC, Trimmomatic [1]); 2) on the side of the genome (Spades [2]); 3) Assembly quality analysis (Quast [3]) 4) genome annotation (Prokka); 5) probable identification of the organism (ProTaxon); 6) search for enzymes (Blast, Prokka); 7) reconstruction of gene networks (ANDSystem [4, 5]. The developed pipeline was used on the data of the collection of microorganisms of the Kurchatov Genomic Center. To date, 2086 genomes have been processed. The average genome length was around 6,000,000 bp. The average number of genes in the genome is 5627. KEGG gene networks and metabolic pathways were reconstructed for all analyzed genomes.

Funding: Work was funded by the Ministry of Science and Higher Education of the Russian Federation project “Kurchatov Center for World-Class Genomic Research” number 075-15-2019-1662 from 2019-10-31.

References

1. Bolger A.M., Lohse M., Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120. doi 10.1093/bioinformatics/btu170
2. Nurk S., Bankevich A., Antipov D. et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol*. 2013;20(10):714-737. doi 10.1089/cmb.2013.0084
3. Gurevich A., Saveliev V., Vyahhi N., Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-10755. doi 10.1093/bioinformatics/btt086
4. Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biol*. 2012;11(3-4):149-161. doi 10.3233/ISB-2012-0449
5. Ivanisenko T.V., Saik O.V., Demenkov P.S. et al. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics*. 2020;21228. doi 10.1186/s12859-020-03557-8

doi 10.18699/bgrs2024-8.4-11

A study of the community relationships between methanotrophs and their satellites using constraint-based modeling approaches

Esembaeva M.A.*, Kulyashov M.A., Kolpakov F.A., Akberdin I.R.
Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia
** esembaevamaryam@gmail.com*

Key words: constraint-based modelling; community modelling; methanotrophs; *Methylococcus capsulatus*; *Cupriavidus necator*; *Escherichia coli*

Motivation and Aim: Methanotrophs are a group of microorganisms that use methane as their primary source of carbon and energy, and they are actively utilized in modern biotechnology [1]. One of the most studied methanotrophs is *Methylococcus capsulatus*, whose growth in monoculture is hindered by the accumulation of repressing byproducts. One of the solutions to this problem is co-cultivation of the methanotroph with other satellite microorganisms, which consume byproducts as substrates for their growth [2]. However, metabolic interactions within such microbial communities remain poorly understood. To assess metabolic interactions in the community, a new systems biology approach is currently applied – the reconstruction of constraint-based models of microbial communities, which allows for the assessment of strain interactions during co-cultivation [3]. Therefore, the aim of this study is to investigate the community model of *Methylococcus capsulatus* and its satellites, using as examples *Escherichia coli* and *Cupriavidus necator*.
Methods and Algorithms: To reconstruct community models, published genome-scale metabolic models of the bacteria *Methylococcus capsulatus* (*i*McBath) [4], *Cupriavidus necator* (*i*CN1361) [5], and two *Escherichia coli* models from the BIGG Models database were used: e_coli_core (http://bigg.ucsd.edu/models/e_coli_core) for the community model development and *i*AF1260 (http://bigg.ucsd.edu/models/iAF1260) for the community model with *M. capsulatus*. Prior to the community model construction, the quality of the *M. capsulatus* and *C. necator* models was assessed using the MEMOTE toolbox [6]. To consider metabolic interactions between two bacterial strains in the community model, transport, exchange reactions and external metabolites for them were added to both models using the Cobrapy library [7]. In the *C. necator* model, reaction and metabolite identifiers were replaced with corresponding BIGG database identifiers to facilitate interactions with the *M. capsulatus* model. The CommunityModel package from the Mewpy library [8] was employed to build the constraint-based community model of the methanotroph with its satellites. The package includes several widely used modules for analyzing microbial community flux balance models. The study was conducted in the BioUML platform [9], and the original program code is presented within a Jupyter Notebook framework (https://shorturl.at/eEOTW).
Results:
Evaluation of the models M. capsulatus, C. necator
Prior to the microbial community model reconstruction, it was necessary to assess the quality of the original models. MEMOTE tool was used for this purpose. The models demonstrated good quality and did not require improvement for further investigation

(detailed reports are available at https://shorturl.at/orFIL, https://shorturl.at/oIOR0 for *i*McBath and *i*CN1361 respectively). At the next step, transport and exchange reactions for acetate were added to the iMcBath model, and the methane uptake rate was increased from 18.46 to 28.1 mmol/gDCW/h^{-1} (gDCW – gram dry cell weight) [10], resulting in an increase in growth rate from 0.178 to 0.27 h^{-1}. Acetate secretion was optimized by selecting acetate exchange reaction as the objective function, with a minimum biomass condition of 0.9 times the maximum. Thus, the acetate exchange rate in the model was amounted to 0.702 mmol/gDCW/h^{-1}.

Pipeline validation for the community model reconstruction

The pipeline for construction of the community model for M. capsulatus and its satellite was tested with selected tools on two combined and simplified (e_coli_core) models of *E. coli*. Acetate was chosen as the main metabolite for interaction between two *E. coli* strains (acetate producer and consumer), as it presumably inhibited the growth rate of *M. capsulatus*, but satellites in mixed culture can use it as a source of carbon. Optimization of acetate secretion by one strain was achieved by adjusting the system to microaerobic cultivation conditions, while for the second strain, the system was adjusted to cultivation conditions with acetate as the carbon source. The community model was constructed using the CommunityModel module from the Mewpy library. As a result, the model yielded the following community metabolites: the interaction between strains occurred via acetate (18.4 mmol/gDCW/h^{-1}), water (17.7 mmol/gDCW/h^{-1}), and hydrogen ions (10.7 mmol/gDCW/h^{-1}). The growth rates of the acetate producer and acetate consumer *E. coli* strains become 0.82 and 0.33 h^{-1}, respectively. This difference can be attributed to the use of different substrates for bacterial growth within the community – acetate being a less efficient substrate compared to glucose that was a carbon source for the acetate producer.

Reconstruction of the community model for Methylococcus capsulatus and Escherichia coli

Following the validation of the pipeline on simplified models, the next step involved constructing a model for the interaction between *M. capsulatus* and *E. coli*. The cultivation medium was chosen based on the microelements required for the models, which are included in the biomass equations. Also, methane was used as a single carbon source for *M. capsulatus* growth. For the *E. coli* model, fluxes were blocked for reactions not present in the community medium, and constraints were set for acetate uptake rate, resulting in a growth rate of 0.46 h^{-1}. Analysis of the resulting community revealed that in the co-cultivation, *M. capsulatus* produced 2.79 mmol/gDCW/h^{-1} of acetate, which was entirely consumed by *E. coli*. Other cross-feeding metabolites are included formaldehyde (1.24 mmol/gDCW/h^{-1}), glycerol (0.0005 mmol/gDCW/h^{-1}), and threonine (1.27 mmol/gDCW/h^{-1}), produced by the methanotroph, and serine (1.27 mmol/gDCW/h^{-1}) produced by *E. coli* (see the Fig. 1).

The growth rates in the community were 0.15 h^{-1} for *M. capsulatus* and 0.077 h^{-1} for *E. coli*. The decrease in the growth rate for the *M. capsulatus* model is attributed to higher acetate production compared to the original model, while the lower growth rate of *E. coli* is associated with the reduced availability of acetate compared to the original model.

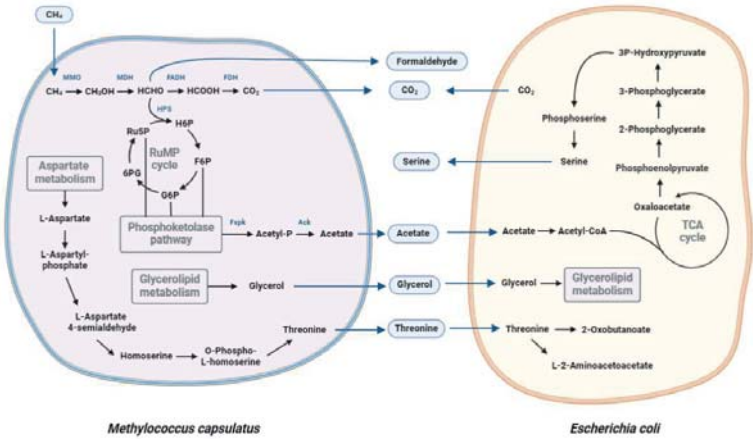


Fig. 1. Interaction between models of *M. capsulatus* (*on the left*) and *E. coli* (*on the right*) after reconstruction of a community model

Conclusion: Using simplified models of *Escherichia coli* as a basis, a pipeline for reconstructed constraint-based metabolic models describing microbial communities was developed. Based on this pipeline was reconstructed a community model for *Methylococcus capsulatus* and *E. coli*, demonstrating their interaction through acetate utilization by satellite strain, the secretion of metabolites such as formaldehyde and serine by the community. Quality assessment of the *M. capsulatus* and *Cupriavidus necator* models was conducted for subsequent model construction and analysis of their interactions.

Funding: The study is supported by the grant of the state program of the "Sirius" Federal Territory "Scientific and technological development of the "Sirius" Federal Territory".

References

1. Kalyuzhnaya M.G. et al. Metabolic engineering in methanotrophic bacteria. *Metab Eng.* 2015;29:142-152. doi 10.1016/j.ymben.2015.03.010
2. Bothe H., Møller Jensen K., Mergel A. et al. Heterotrophic bacteria growing in association with *Methylococcus capsulatus* (Bath) in a single cell protein production process. *Appl Microbiol Biotechnol.* 2002;59:33-39. doi 10.1007/s00253-002-0964-1
3. Lee H. et al. Syntrophic co-culture of a methanotroph and heterotroph for the efficient conversion of methane to mevalonate. *Metab Eng.* 2021;67:285-292. doi 10.1016/j.ymben.2021.07.008
4. Lieven C. et al. A Genome-scale metabolic model for methylococcus capsulatus (bath) suggests reduced efficiency electron transfer to the particulate methane monooxygenase. *Front Microbiol.* 2018;9:2947. doi 10.3389/fmicb.2018.02947
5. Pearcy N. et al. A genome-scale metabolic model of Cupriavidus necator H16 integrated with TraDIS and transcriptomic data reveals metabolic insights for biotechnological applications. *PLoS Comput Biol.* 18(5):e1010106). doi 10.1371/journal.pcbi.1010106
6. Lieven C., Beber M.E., Olivier B.G. et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol.* 2020;38(3):272-276. doi 10.1038/s41587-020-0446-y
7. Ebrahim A., Lerman J.A., Palsson B.O. et al. COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Syst Biol.* 2013;7:74. doi 10.1186/1752-0509-7-74
8. Pereira V., Cruz F., Rocha M. MEWpy: a computational strain optimization workbench in Python. *Bioinformatics*. 2021;37(16):2494-2496. doi 10.1093/bioinformatics/btab013
9. Kolpakov F. et al. BioUML-towards a universal research platform. *Nucleic Acids Res.* 2022;50(W1):W124-W131. doi 10.1093/nar/gkac286
10. Gupta A. et al. Genome-scale metabolic reconstruction and metabolic versatility of an obligate methanotroph *Methylococcus capsulatus* str. Bath. *PeerJ*. 2019;7:e6685. doi 10.7717/peerj.6685

doi 10.18699/bgrs2024-8.4-12

A software package to assess the metabolic potential of mutant strains of the *Corynebacterium glutamicum*

Kazantsev F.V.[1, 2, 3]*, Trofimova M.F.[1, 2], Sheremetieva M.E.[4], Khlebodarova T.M.[1, 3], Matushkin Yu.G.[1, 2, 3], Lashin S.A.[1, 2, 3]

[1] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[4] *National Research Center "Kurchatov Institute", Kurchatov Genomic Center, Moscow, Russia*

* *kazfdr@bionet.nsc.ru*

Key words: Mathematical modelling, FBA, flux balance analysis, high-performance computing.

Motivation and Aim: In the bacterial biotechnology industry, the process of perspective strains searching is never ends. In this search, experimental approaches generate arrays of genomic, metabolomics and other types of data for many bacterial strains. Analysis of such data deals with approaches of bioinformatics, systems biology, mathematics and computational modeling. It allows one to propose a scheme of bacterial metabolic network functioning consistent with the observed data, and to suggest possible scenarios for biotechnological and/or genomic manipulations to optimize bacterial metabolism. In the study, we propose a semi-automated approach and its software implementation to analyze the evaluation of the metabolic potential of mutant strains and its validation on the example of *C.glutamicum* bacterial strains.

Methods and Algorithms: Software package is developed with programming language Python, and libraries cobraPy (cobrapy.readthedocs.io) and Pandas. Metabolic maps creation and visual interpretation of simulation experiments are performed with Escher tool (escher.readthedocs.io).

Results: To organize the computational flow for the analysis of mutation variants of the target strain, a software package was developed that implements a series of computational protocols. The computational protocols are presented in the form of Jupyter notebooks. This solution does not require a researcher to have a high level of proficiency in programming languages, and allows modifications of the work scenario if necessary. The program accepts a ready-to-use whole-genome flux model of bacteria in cobraPy format as input. The following functions are available: analysis of growth conditions on culture media; assessment of the potential of mutant strains. Implementations of these functions requires screening the parameters for hundreds of metabolic reactions within the computational experiment with further verification of acceptable solutions, which requires running the analysis on high-performance computing clusters.

Funding: The study is supported by the Kurchatov Genomic Center No. 075-15-2019-1659 and No. 075-15-2019-1662.

doi 10.18699/bgrs2024-8.4-13

Quantifying ecological functional groups in the human gut microbiome: metagenomics meets trait-based ecology

Klimenko A.I.[1, 2]*, Kropochev A.I.[1], Lashin S.A.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *klimenko@bionet.nsc.ru*

Key words: trait-based ecology; ecological functional groups; metagenomics; human gut microbiome

Motivation and Aim: The human gut microbiota plays a crucial role in digestive function, metabolism of short-chain fatty acids, vitamins and other compounds, immune function and maintenance of body health. Yet most software tools for functional analysis of metagenomic and metatranscriptomic data from the gut microbiota assess a wide range of potentially active functions and do not target specific processes that are key to a given ecosystem. Moreover, since generally functional traits of microorganisms have a limited phylogenetic conservatism trait-based ecology appears to be a well-suited approach to describe such complex ecosystems operating in terms of interactions between the ecological functional groups (EFGs) identified on the basis of ecologically relevant traits rather than taxonomic units. In this study we use specific trait-determining genetic features (TDGFs) to distinguish EFGs.

Methods and Algorithms: The program for the analysis of the ecological structure of the symbiotic human gut microbiota (EcoStructSHGM) is implemented in Python as a bioinformatics pipeline. We use Kallisto for TDGF quantification, jellyfish for k-mer quantification, bowtie2 and fastp for quality and contamination control as well as the algorithms that we have developed specifically for quantifying EFGs. The UniProtKB database has been used to create amino-acid catalogue of human gut microbiota TDGFs.

Results: We have developed a method for bioinformatic quantification of ecological functional groups in the human gut microbiota covering such processes as acetogenesis, butyrate production, sulfate reduction and mucin degradation, which employs available metagenomic or metatranscriptomic data to quantify the abundance or activity of relevant EFGs. The developed method was verified using publicly available data from the study of the effect of diet on mucin layer stability and compared with HumanN 3 software for the functional metagenomics analysis showing high consistency and recall of our method. We have also created amino-acid catalogue of human gut microbiota TDGFs, which allows to use EcoStructSHGM for analysing taxonomically diverse microbial communities.

Conclusion: The developed method for quantifying ecological functional groups in the human gut microbiome based on metagenomics data takes advantage of trait-based ecology approach to identify ecological functional groups in samples by their specific trait-determining genetic features, which allows to focus a functional metagenomics analysis on the processes, which are highly relevant for an ecosystem under investigation.

Funding: The study is supported by the Budget Project FWNR-2022-0006.

doi 10.18699/bgrs2024-8.4-15

Metabolic modeling of the production of a recombinant protein by methanotrophs

Kulyashov M.A.[1]*, Richard H.[2], Kolmykov S.K.[1], Sokolova T.S.[1], Khlebodarova T.M.[1], Kalyuzhnaya M.G.[2], Akberdin I.R.[1]

[1] *Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia*

[2] *San Diego State University, San Diego, USA*

* *kulyashov.ma@talantiuspeh.ru*

Key words: constrained-based modelling; methanotrophs; recombinant protein

Motivation and Aim: Methanotrophic bacteria are widely used in advanced biotechnology, beginning from the production of single cell protein (SCP), and ending with different cell metabolites: succinate, polyhydroxyalkanoates (PHAs), ectoine, methanol and others. Main advantage of their application in biotechnology is using a low-cost and available source – methane. One of these organisms widely studied among methanotrophs is *Methylotuvimicrobium alcaliphilum* 20Z^R, for which the production of recombinant proteins is the new biotechnology field of its application in a few last year's. Moreover, a genome-scale metabolic (GSM) model verified by a few omics experimental datasets was developed for the strain [1, 2]. The main goal of the study is an optimization of the recombinant protein production by *Methylotuvimicrobium alcaliphilum* 20Z^R using constraint-based modeling approaches.

Methods and Algorithms: The following proteins were used as target recombinant proteins: sfGFP protein, the nucleotide sequence of which was provided by Prof. Marina Kalyuzhnaya (San Diego California State University, USA); the amino acid sequence of the bovine β-casein protein (https://www.uniprot.org/uniprot/P02666), which was taken from the Uniprot database [3], was also used. Two marker proteins were also analyzed: spyCatcher and spyTAG which have 114 and 17 amino acids length, respectively. Their sequences as well as an estimation of ATP consumption for the production of recombinant protein were also provided based on the experiments in Kalyuzhnaya's lab. To run the algorithm, the growth medium on methane specified in the original model was used, with a methane consumption rate of 11.7 mmol*gDCW^{-1}*h^{-1} (gDCW – gram of dry biomass), with the exception of the context-specific model, in which a methane consumption rate of 2.85 mmol was used mmol*gDCW^{-1}*h^{-1} in accordance with experimental measurements.

Model modification and extension was made in the BioUML platform [4] using python libraries Biopython [5] and Cobrapy [6]. All original code and developed modules are available on the BioUML platform – https://shorturl.at/cmG47.

Co-optimization of recombinant protein production in an extended model was made using the Optflux tool [7] with WYELD evolutionary function, which gives an opportunity to fix the lower bound for the biomass equation, by taking in account the biomass fraction which is determined by the user.

Results: The first step was to write a program code in Python, which, based on the sequence of the recombinant protein, would create a pseudo-equation for its synthesis in a GSM model, and also add the set of reactions for its transport and exchange, taking

into account energy costs. Using the Biopython and Cobrapy libraries, the module was written. It allows a user to pass not only the amino acid sequence of a protein, but also the nucleotide sequence, which is converted into an amino acid, and then counts each type of amino acid, generates an identifier for it in the model, creates a pseudo-equation for the synthesis of a recombinant protein, taking into account the procedure for calculating stoichiometric coefficients and energy costs for the production of the corresponding protein. After generation of the pseudo-equation for protein synthesis, the transport reaction of its export from the cell is added to the model, taking into account ATP costs for its transport. In addition, this new BioUML module also enables to immediately co-optimize the production of recombinant protein using the evolutionary algorithm via the OptKnock algorithm implemented in the Mewpy library [8]. The original written code provides a feasibility to set various parameters for co-optimization, including the selection of environmental parameters of interest for simulation of the cell culture growth. The co-optimization can be conducted by means of 2 evolutionary functions, BPCY and WYIELD or via the use of both simultaneously. The written code implements options for running these algorithms for one target function, which expands the choice of algorithms to search for potential genetic modifications. A graphical interface of the code was also written that makes it easy to use for researchers without a programming background.

The previously constructed *i*IA409 GSM model [2] has been extended to integrate and account for the synthesis of recombinant proteins using the developed module. Furthermore, several models have been built in order to co-optimize and predict the production of different recombinant proteins in the combination with various marker proteins used simultaneously in the growth of the cell culture. Analysis of co-optimization results of these modified models depending on the used combination of both recombinant and marker proteins, enabled to identify potential genetic modifications facilitating the production of the corresponding protein simultaneously with the growth of the cell culture at various acceptable minimum values of the biomass obtained for wild type simulation.

Thus, three main directions of genetic modifications have been revealed both for the GFP protein and for the β-casein protein. The first set of modifications include modifications associated with a *direct change in the activity of TCA cycle reactions***: an increase in the activity** of citrate synthase, fumarate reductase, and malate dehydrogenase. The efficiency of these modifications can be experimentally verified. Another direction of modifications which is of interest, are reactions associated with the *redistribution of the carbon flux* in the cell due to the **activation** of the H_4MPT pathway, as well as a **decrease in the activity** of formate dehydrogenase, which will reduce the conversion of carbon into CO_2. Last one direction for modifications **is the increase of ATP production**: increase the activity of ATP synthase, increase the availability of H^+ in the cell, which would also lead to an increase of ATP production by the cell. However, this type of modifications is more challenging for experimental verification.

It is worth to note that the list of proposed modifications depends on the amino acid composition of the heterologous protein sequence. It indicates the developed program module with the implemented algorithm to search for potential modifications for various recombinant proteins takes into account the specificity of amino acids profile.

Conclusion: The Python-based module in the BioUML platform, which gives an opportunity to modify GSM model of methanotrophic organisms to account for synthesis of recombinant protein by the cell was developed The algorithm implemented in the

module considers the energy costs for protein synthesis and transport. Moreover, the module was applied to the original GSM model, *i*IA409 for *Methylotuvimicrobium alcaliphilum* 20Z^R, and allowed us to identify a set of potential genetic modifications contributing to the enhancement of recombinant protein production by the methanotroph taking into account the amino acids profile of the required protein.

Funding: The study was financially supported by the Russian Science Foundation (project № 23-24-00606, https://rscf.ru/en/project/23-24-00606/).

References

1. Akberdin I.R. et al. Methane utilization in Methylomicrobium alcaliphilum 20Z^R: a systems approach. *Sci Rep*. 2018;8(1):2512. doi 10.1038/s41598-018-20574-z
2. Akberdin I.R. et al. Rare earth elements alter redox balance in *Methylomicrobium alcaliphilum* 20Z^R. *Front Microbiol*. 2018;9:2735. doi 10.3389/fmicb.2018.02735
3. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*. 2023;51(D1):D523-D531. doi 10.1093/nar/gkac1052
4. Kolpakov F. et al. BioUML-towards a universal research platform. *Nucleic Acids Res*. 2022;50(W1):W124-W131. doi 10.1093/nar/gkac286
5. Cock P.J. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009;25(11):1422-1423. doi 10.1093/bioinformatics/btp163
6. Ebrahim A., Lerman J.A., Palsson B.O., Hyduke D.R. COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Syst Biol*. 2013;7:74. doi 10.1186/1752-0509-7-74
7. Vilaça P., Maia P., Giesteira H., Rocha I., Rocha M. Analyzing and designing cell factories with OptFlux. *Methods Mol Biol*. 2018;1716:37-76. doi 10.1007/978-1-4939-7528-0_2
8. Pereira V., Cruz F., Rocha M. MEWpy: a computational strain optimization workbench in Python. *Bioinformatics*. 2021;37(16):2494-2496. doi 10.1093/bioinformatics/btab013

doi 10.18699/bgrs2024-8.4-16

Study of the bacterial community of the gastric mucosa in patients with gastro-duodenal region diseases using metatranscriptomic analysis

Kupriyanova E.*, Abdulkhakov S., Markelova M., Grigoryeva T.
Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia
** fewrandomletters@mail.ru*

Key words: gastric microbiota; metatranscriptome; sequencing; *H. pylori*

Motivation and Aim: It has now been proven that *H. pylori* infection can lead to chronic inflammatory changes in the gastric mucosa, and subsequently to the development of precancerous changes in the mucous membrane [1]. However, other types of bacteria, in addition to *H. pylori*, can provoke the development of chronic inflammatory changes in the gastric mucosa, and even contribute to the development of gastric cancer [2]. But the mutual influence of *H. pylori* and other representatives of the gastric microbiota remains poorly understood. Taking into account the morphological and functional characteristics of different parts of the stomach, including those caused by the production of hydrochloric acid, it is obvious that the microbial composition in different parts of the stomach may differ.

In the vast majority of studies, the gastric microbiota has been studied at the bacterial DNA level using sequencing of 16S rRNA gene libraries, which include DNA from dead or destroyed cells. In this regard, the purpose of this study was to analyze the resident, metabolically active representatives of the microbiota in the two main parts of the stomach – the body and the antrum – using sequencing of RNA-derived libraries.

Methods and Algorithms: We obtained 120 paired gastric mucosal biopsies (one from the antrum and one from the corpus, respectively) from 60 patients who underwent esophagogastoduodenoscopy (EGD) for suspected gastroduodenal disease. Total RNA from each sample was extracted with TRIzol reagent. The RNA fraction was then used to synthesize cDNA. Amplicon libraries of the V3-V4 variable region of the 16S rRNA gene were sequenced on the Illumina MiSeq platform.

Results: In the majority of *H. pylori*-positive patients, a significant (up to 90 %) dominance of the *Helicobacter* genus was found. Bacteria of the *Streptococcus* genus, in particular the species *Streptococcus vestibularis*, as well as the *Prevotella* and *Alloprevotella* genera, were represented to a lesser extent. Previous studies examining the composition of the gastric microbiota in patients with chronic gastritis showed that there is a significant negative correlation between the presence of *H. pylori* and the proportion of the *Streptococcus* genus: the presence of *H. pylori* inhibits the growth of *Streptococcus*, and in *H. pylori*-negative patients their number increases [3]. Indeed, in the absence of *H. pylori* infection, we found an increase in the representation of *Streptococcus* by more than 2 times, as well as the presence in a larger number of the *Prevotella*, *Alloprevotella Gemella*, *Haemophilus*, *Staphylococcus*, *Sphingomonas* genera, which, according to the literature, can be considered as representatives of the normal gastric microbiota [4].

There is also a difference between the bacterial composition of the antrum and body of the stomach, which is especially important for *H. pylori*-positive patients. For example, the abundance of *Helicobacter* species in individual samples was 25, 35, and 36 % in the gastric antrum and 87, 81, and 72 % in the corpus, respectively, calling into question the rationale for taking biopsies only from the antrum to confirm the diagnosis of *H. pylori* infection. Also, bacteria from the *Lautropia*, *Enhydrobacter*, *Rikenellaceae*, *Acinetobacter*, *Leptotrichia*, and *Capnocytophaga* genera were found in low numbers in many samples, so they likely represent a minor, poorly studied part of the gastric microbiota.

Based on the data obtained, it can be assumed that *H. pylori* tends to predominate in the stomach of *H. pylori*-positive patients, especially when analyzing the body of the stomach compared to the antrum. *H. pylori* exhibits antagonistic activity against bacteria of the *Streptococcus* genus; in *H. pylori*-negative patients, the number of *Streptococcus* increases significantly. The use of gastric corpus biopsy is likely to have greater prognostic value in the diagnosis of *H. pylori* infection. Also, the use of RNA, compared to classical DNA-based methods, allows the identification of a greater diversity of bacterial genera.

Funding: This research was funded by the subsidy allocated to the Kazan Federal University for the state research assignment (No. FZSM-2023-0013) and the Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030).

References

1. Uemura N., Okamoto S., Yamamoto S. et al. Helicobacter pylori infection and the development of gastric cancer. *N Engl J Med.* 2001;345(11):784-789. doi 10.1056/NEJMoa001999
2. Bik E.M., Eckburg P.B., Gill S.R., Nelson K.E. et al. Molecular analysis of the bacterial microbiota in the human stomach. *Proc Natl Acad Sci USA.* 2006;103(3):732-737. doi 10.1073/pnas. 0506655103
3. Parsons B.N., Ijaz U.Z., D'Amore R., Burkitt M.D., Eccles R., Lenzi L. et al. Comparison of the human gastric microbiota in hypochlorhydric states arising as a result of Helicobacter pylori-induced atrophic gastritis, autoimmune atrophic gastritis and proton pump inhibitor use. *PLoS Pathog*. 2017;13(11):e1006653. doi 10.1371/journal.ppat.1006653
4. Rajilic-Stojanovic M., Figueiredo C., Smet A. et al. Systematic review: gastric microbiota in health and disease. *Aliment Pharmacol Ther*. 2020;51(6):582-602. doi 10.1111/apt.15650

doi 10.18699/bgrs2024-8.4-17

Mathematical modeling of the regulation of NCgl2816-lldD, brnEF, vanABK and mtlTD operons in *Corynebacterium glutamicum*

Lakhova T.[1, 2, 3]*, Kazantsev F.[1, 2, 3], Khlebodarova T.[1, 2], Matushkin Yu.[2, 3], Lashin S.[1, 2, 3]

[1] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *tlakhova@bionet.nsc.ru*

Key words: bacterial transcription regulation; *Corynebacterium glutamicum*; transcription factors

Motivation and Aim: *Corynebacterium glutamicum* is one of the model organisms and one of the most demanded in the field of industrial biotechnology. Various strains of *C. glutamicum* are used to produce more than 70 chemicals [1]. An important rule is the influence of transcription factors on the level of gene expression. However, there is still no complete network of transcriptional regulation in bacteria. Such a network would allow scientists to more easily determine where to manipulate bacteria in order to increase the yield of the target product. Mathematical models are created to help understand how transcription factors can affect the level of gene expression. That is, to determine their contribution to the regulation of any operons. This allows us to consider target genes and/or transcription factors in terms of point selection of candidates that need to be tested by modeling before genetic modifications can be made in bacteria. In this study, we constructed and analyzed mathematical models of the following operons: *NCgl2816-lldD*, *brnEF*, *vanABK*, and *mtlTD*. The *NCgl2816-lldD* operon is responsible for lactate utilization, *brnEF* is a transporter of branched-chain amino acids or L-methionine, *vanABK*, whose genes are involved in vanilate utilization, and *mtlTD* is responsible for mannitol catabolism in the bacterium. These operons were chosen, firstly, because experimental data were found for them, and secondly, to demonstrate the independence of the choice of any operon, as study would be similar.

Methods and Algorithms: To create a model of transcriptional regulation of an operon, we need to know the following information: transcription factors (TFs) and their binding sites located on DNA, the number of TFs regulating a given operon, and the type of regulation. For the selected operons, we found this information about the participants of regulation from the articles. We then generate frame models from this data using Operon_Equations [2]. The Operon_Equations reconstructs the mathematical model in differential equations in terms of generalized Hill functions. However, such models may exhibit general behavior and have unknown parameters. We used WebPlotDigitizer (version 4.6) [3] to extract experimental data from article figures to verify our models and determine the parameters for each model. The Scipy package was used to fit the parameters.

Results: The following information was obtained from the articles: *NCgl2816-lldD*, *vanABK* and *mtlTD* are characterized by repressive effects of TFs, while *brnEF* is characterized by activation effects. Moreover, an effector is required for each of them.

No such study has been performed for RamA as a TF of *NCgl2816-lldD*, but some experiments indicate that RamA is a TFs relative to *NCgl2816-lldD*, so we decided to include it in our study. We created mathematical models for four operons: *NCgl2816-lldD*, *brnEF*, *vanABK*, and *mtlTD*. For each of them, the parameters were matched according to the selected values within the life capacity of the bacteria. Verification of these models has been carried out.

Funding: The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics SB RAS No. 075-15-2019-1662.

References

1 Becker J., Rohles C.M., Wittmann C. Metabolically engineered corynebacterium glutamicum for bio-based production of chemicals, fuels, materials, and healthcare products. *Metab Eng.* 2018;50:122-141. doi 10.1016/j.ymben.2018.07.008

2 Lakhova T.N., Kazantsev F.V., Mukhin A.M., Kolchanov N.A., Matushkin Y.G., Lashin S.A. Algorithm for the reconstruction of mathematical frame models of bacterial transcription regulation. *Mathematics*. 2022;10:4480. doi 10.3390/math10234480

3 Rohatgi A. WebPlotDigitizer Available online: https://automeris.io/WebPlotDigitizer.html (accessed on 4 October 2023)

doi 10.18699/bgrs2024-8.4-18

Insights into the natural activation mechanism of the CRISPR-Cas immune system in *Escherichia coli* from a computational modeling analysis

Rodic A.[1]*, Tumbas M.[1], Kondev J.[2], Djordjevic M.R.[3], Djordjevic M.J.[1]

[1] *Faculty of Biology, University of Belgrade, Belgrade, Serbia*

[2] *Martin A. Fisher School of Physics, Brandeis University, Waltham, MA, USA*

[3] *Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia*

* *andjela.rodic@bio.bg.ac.rs*

Key words: CRISPR-Cas derepression; H-NS; modeling expression dynamics, machine learning

Motivation and Aim: CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR-Associated proteins) systems provide bacteria with defense against viruses by utilizing spacers, fragments of viral DNA stored within the CRISPR array, a locus in a bacterial genome. These spacers are transcribed and processed into crRNAs, which guide Cas proteins to target and eliminate matching viral DNA sequences. Despite the extensive biotechnological applications of CRISPR-Cas in gene editing, its natural functioning within bacteria remains incompletely understood. Particularly in *Escherichia coli*, the most extensively studied system, CRISPR-Cas activity is typically silenced under experimental conditions. This silencing is attributed to the repression of promoters governing the expression of *cas* genes and the CRISPR array, facilitated by cooperative binding of the global regulatory protein H-NS (Histone-like Nucleoid-Structuring). Moreover, a fraction of H-NS proteins might become sequestered by viral DNA if it possesses a higher AT content compared to bacterial DNA [1]. Additionally, the pleiotropic transcriptional regulator, LeuO, whose expression is modulated by BglJ (encoded by the β-glucoside operon), can alleviate H-NS-mediated repression [2]. This study aimed to investigate whether a modest reduction in the cellular H-NS level could initiate a positive feedback loop, activating the CRISPR-Cas system through BglJ and LeuO proteins, thereby leading to rapid crRNA generation for effective cellular defense against fast-replicating bacteriophages.

Methods and Algorithms: We developed a mathematical model to predict the temporal dynamics of crRNA levels during the activation of the CRISPR-Cas system upon infection of a bacterial cell by foreign DNA. This model, utilizing principles of statistical thermodynamics and non-linear dynamics, incorporates known regulatory features of the system as well as our proposed mechanism of combined activation, by: 1) derepression of the *cas* genes promoter due to sequestration of H-NS by foreign DNA, and 2) activation of *cas* genes expression through positive feedback performed by LeuO and BglJ-RcsB. We systematically searched for parameter combinations at which the increase in the crRNA amount, indicative of the activation of the CRISPR-Cas system, can occur during the first 30 min upon entry of the foreign, AT-rich DNA. To identify the parameters exerting the most significant influence on eliciting this response from the system, we applied the Random Forest machine learning technique. Additionally, we conducted a bioinformatics analysis where 16,388 viruses were matched to their respective host bacteria harboring CRISPR-Cas using the Virus-Host DB database and

tools for identifying CRISPR arrays in bacterial genomes. For each bacterial genus, we assessed the disparity in genomic AT content between bacterial and viral DNA.
Results: Our analysis revealed a consistent, small elevation in AT content within viral genomes compared to host genomes, supporting a potential reduction in available H-NS levels for transcriptional repression upon bacteriophage DNA introduction into a cell. The Random Forest analysis showed that the most important parameters are: 1) the baseline cellular H-NS level, 2) the change in accessible H-NS concentration as a consequence of its interaction with foreign DNA, and 3) the degree of cooperativity associated with H-NS binding to promoter DNA. Our findings indicate that substantial reductions in the available H-NS levels can readily induce crRNA expression. On the other hand, achieving timely crRNA generation from a minor reduction in H-NS levels necessitates high H-NS binding cooperativity and a baseline H-NS level close to its equilibrium dissociation constant for binding to DNA.
Conclusion: Our modeling approach suggests that CRISPR-Cas systems possess the capability to mount a rapid response in the face of fast-replicating bacteriophages, generating sufficient crRNAs for effective defense. The identified kinetic features essential for this response, which should be the focus of future experimental work, contribute to understanding the intrinsic behavior of the system, crucial for the development of biotechnological tools leveraging bacterial hosts for compound production and the design of antiviral platforms.
Funding: This work is supported by the Science Fund of the Republic of Serbia (projects No. 7750294, q-bioBDS, and No. 6417603, CRISPR modelling).

References

1. Pul U., Wurm R., Arslan Z., Geissen R., Hofmann N., Wagner R. Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Mol Microbiol*. 2010;75(6):1495-1512. doi 10.1111/j.1365-2958.2010.07073.x
2. Westra E.R., Pul U. et al.H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. *Mol Microbiol*. 2010;77(6):1380-1393. doi 10.1111/j.1365-2958.2010.07315.x

doi 10.18699/bgrs2024-8.4-19

Reconstruction of regulons for *Methylotuvimicrobium alcaliphilum 20Z^R* based on an analysis of a bunch of transcriptomic data

Sokolova T.S.[1]*, Kolmykov S.K.[1], Kulyashov M.A.[1], Khlebodarova T.M.[1], Kalyuzhnaya M.G.[2], Akberdin I.R.[1]

[1] *Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia*

[2] *Biology Department and Viral Information Institute, San Diego State University, San Diego, USA*

** lynxl@list.ru*

Key words: RNA-seq; transcription factor; regulon reconstruction; methanotroph; *Methylotuvimicrobium alcaliphilum 20Z^R*

Motivation and Aim: The study of methanotrophic bacteria represents a "hot" research field in microbial biotechnology, as these organisms have the potential to utilize methane as a carbon source for the production of useful target and value-added compounds. Complex interactions between transcription factors (TFs) and target genes are responsible for various molecular and phenotypic features of bacteria, including adaptation to changes in environmental and/or cultivation conditions. Despite the long-term history of methanotrophs' investigation, both quantitative/qualitative data and knowledge of such interactions are very limited. However, the assumption that regulatory interactions between TFs and target genes are conservative among bacteria allows us to use the accumulated omics data to reconstruct the regulatory structure in closely related species. Herein, we employ the data on expression profiles for two closely related methanotrophs, *Methylotuvimicrobium alcaliphilum 20Z^R* and *Methylotuvimicrobium buryatense 5GB1*, to construct the first version of regulons in methane-oxidizing bacteria. This finding sheds light on transcriptional modules regulating molecular machinery in methanotrophs at the first time, to the best of our knowledge. Furthermore, the outcome of this study can be essential to pave the way in the fruitful development of biotechnological applications of methane-oxidizing microorganisms.

Methods and Algorithms: Transcriptomic datasets, both unpublished original and previously published [GSE51145, GSE101981, GSE253414, GSE125909, GSE221011] were used for the regulons structure. These datasets were obtained under growth conditions, when methane or methanol are used as a carbon source, and metal ions, including copper, tungsten, calcium, and lanthanum are harnessed as potential regulators of enzymes' activities involved in the metabolism of these substrates. Gene co-expression networks (GCNs) for *M. alcaliphilum 20Z^R* and *M. buryatense 5GB1* were built using the GENIE3 algorithm [1]. The obtained sets of top gene pairs were divided into 30 clusters using the ward.D2 hierarchical clustering algorithm. Functional enrichment analysis of the clusters was carried out using homebrew R-script. The Operon-mapper web service was used to predict the operon structure of *M. alcaliphilum 20Z^R* [2]. To identify potential regulators for *M. alcaliphilum 20Z^R* we use databases with computationally predicted TFs: deepTF [3], P2TF [4] and ENTRAF DB [5]. TFBS

enrichment analysis for differentially expressed clusters was performed using the MEME suite [6].

Results: The resulting GCN, containing 30 clusters, consisted of 1,241 genes, 78 of which were putative TFs supported by at least one of the sources. To functionally annotate the resulting clusters we conducted enrichment analysis based on KEGG pathways, COG categories and GO terms. Half of the clusters have more than 1 enriched KEGG pathway and 7 clusters do not show any pathway enrichment. 24 clusters demonstrate enrichment in at least one COG functional category, 11 of which are related to only one category. Only 2 clusters show enrichment in one GO term, while 3 clusters do not show any significant enrichment of the GO terms. Thus, the remaining 25 clusters have more than one enriched GO term. We analyzed gene expression levels in the clusters for 3 conditions: CH_4 limitation, O_2 limitation and Ca-La switch. Based on the average gene expression level in the cluster, we divided the clusters into two groups: up and down regulated; the clusters that did not show significant changes in average gene expression levels were ignored. As a result, several known co-regulated groups of genes have been identified, such as the *mxaF* and *mxaI* genes encoding Ca-dependent methanol dehydrogenase, and related genes *mxaR* and *mxaJ*, which have reduced expression levels in the presence of lanthanides [7]. Several new clusters of genes controlled by one or more potential TFs have also been discovered. As an example, we observed that two co-expressed clusters (clusters 14 and 24) had significantly reduced expression under methane- and oxygen-limited growth conditions. Interestingly, these clusters could be combined into one cluster by two potential TFs, MEALZ_RS12170 (feoC) and MEALZ_RS18500 (Fig. 1).

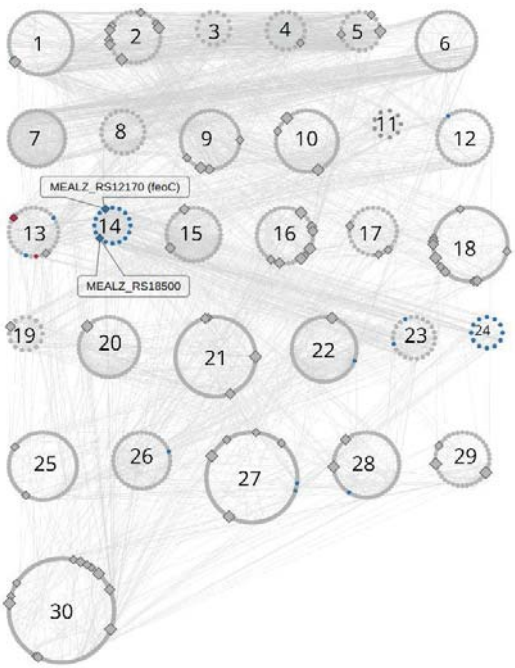


Fig. 1. Down-expressed clusters 14 and 24 in low CH_4 growth condition. Blue color of both circles and diamonds in a cluster means down-expressed genes and potential TFs (MEALZ_RS12170 (feoC) and MEALZ_RS18500) correspondingly. Genes in the cluster are associated with the iron metabolism

We hypothesize that these genes encode true TFs whose expression and genes in their operons are associated with the iron metabolism or reduction in response to methane and/or oxygen limitation. In addition to the clusters identified and described above, a cluster of genes mainly related to nitrogen metabolism transport and regulation has been shown to have reduced expression levels under hypoxia conditions. Their only co-expressed potential TF, CRP, is a regulator of carbon source utilization gene expression in the absence of glucose. It also regulates the expression of GlnG, a known TF mainly regulating nitrogen transport and metabolism in the bacterial cell. In order to confirm the potential regulatory role of revealed transcription factors we performed a search of regulatory signals (motifs) in the upstream of promoters for genes highly co-expressed with the potential TF in a bunch of analyzed transcriptomic profiles. Furthermore, we alternatively explore the regulatory motifs via an interrogation of orthologous genes encoding those genes in closely related methanotrophic strains.

The source code and all data for the whole analysis are available on a GitHub repository https://github.com/Lynxlazy/20ZR-Regulons.

Funding: The study was financially supported by the Russian Science Foundation (project No. 23-24-00606, https://rscf.ru/en/project/23-24-00606/).

References

1. Huynh-Thu V.A. et al. Inferring regulatory networks from expression data using tree-based methods. *PLoS One*. 2010;5(9):e12776. doi 10.1371/journal.pone.0012776
2. Taboada B. et al. Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*. 2018;34(23):4118-4120. doi 10.1093/bioinformatics/bty496
3. Bao X.R. et al. DeepTF: Accurate prediction of transcription factor binding sites by combining multi-scale convolution and long short-term memory neural network. In: Intelligence Science and Big Data Engineering. Big Data and Machine Learning. IScIDE 2019. Lecture Notes in Computer Science. Vol. 11936. Springer, 2019;126-138. doi 10.1007/978-3-030-36204-1_10
4. Ortet P. et al. P2TF: a comprehensive resource for analysis of prokaryotic transcription factors. *BMC Genomics*. 2012;13:628. doi 10.1186/1471-2164-13-628
5. Ledesma L. et al. Prediction of DNA-binding transcription factors in bacteria and archaea genomes. In: Prokaryotic Gene Regulation. Methods in Molecular Biology. Vol. 2516. New York: Humana, 2022;103-112. doi 10.1007/978-1-0716-2413-5_7
6. Bailey T.L., Johnson J., Grant C.E., Noble W.S. The MEME Suite. *Nucleic Acids Res*. 2015;43(W1):W39-W49. doi 10.1093/nar/gkv416
7. Akberdin I.R. Rare earth elements alter redox balance in *Methylomicrobium alcaliphilum 20Z^R*. *Front Microbiol*. 2018;9:2735

doi 10.18699/bgrs2024-8.4-20

Bioinformatics analysis of the biophysical determinants of CRISPR array adaptation mechanism

Tumbas M.*, Đorđević M.
Quantitative Biology Group, Faculty of Biology, University of Belgrade, Belgrade, Serbia
** marko.tumbas@bio.bg.ac.rs*

Key words: CRISPR-cas; melting energy; extreme value distribution

Motivation and Aim: CRISPR-cas systems are incredibly diverse and currently are classified into six major types and over 30 subtypes [1]. Apart from their role in adaptive immunity, some of the CRISPR-cas subtypes are also involved in host gene regulation and collateral damage, leading to bacteriostatic or lethal outcomes for the host. CRISPR array spacers direct and influence canonical and non-canonical functions of the CRISPR-Cas system together with subtype Cas proteins. A better understanding of spacer adaptation mechanisms is crucial for uncovering the intricacies of an evolutionary arms race between prokaryotes and phages.
Methods and Algorithms: Bacterial and viral genomes were retrieved using NCBI Datasets API [2]. CRISPRidentify and CRISPRcasIdentifier tools were used for CRISPR array, Cas genes detection, and subtyping [3–4]. Viral genomes were mapped to their hosts using the latest version of the Virus-Host DB [5]. Mapping was performed on the genus level of the hosts' phylogenetic trees. Gumbel extreme value distribution was used to determine the statistical significance of each spacer Smith-Waterman alignment score. Mobile genetic elements (MGE) and prophage predictions were obtained for all complete bacterial genomes using the VRprofile2 tool [6].
Results: We present a large-scale analysis of CRISPR array spacers from 31,845 complete bacterial genomes. Differences in melting energy and GC content between identified spacers, origin bacterial genomes, and infecting bacteriophages were explored for different CRISPR-cas subtypes and bacterial genera. Significant GC content differences can be observed between bacterial genomes and their spacers and between spacers and infecting phage genomes. Spacers from the extremes of the GC content distribution were aligned to bacterial and infecting phage genomes to determine their origin. We obtain that high GC content spacers have better alignments to phage genomes than low GC spacers and that high and low GC spacers show better alignments to origin bacterial genomes than infecting phage genomes. We further tested if spacer alignment locations to origin bacterial genomes correspond to predicted MGE and Prophage locations and if alignment scores between those spacer groups differ. We found that low GC spacers do not preferentially target integrated prophage or MGE locations over the rest of the bacterial genomes, while high GC spacers tend to target integrated prophage sequences over MGE and remaining bacterial genome sequences.
Conclusion: The GC content of the spacers was smaller than the GC content of the source bacterial genome but larger than the infecting viral genome. This observation aligns with the hypothesis that the majority of CRISPR spacers were adapted from the bacteriophage genomes and serve a canonical function. Alignments of the spacers from GC-rich distribution tails have shown their preferential targeting of host genomes. This further supports the hypothesis that GC-rich spacers originated from the bacterial genome and

have a non-canonical function. The alignment of GC-rich spacers to predicted prophage genomes confirms their viral origin and further supports the hypothesis of GC-rich spacers' preferential adaptation to CRISPR arrays.

Funding: The study is supported by the q-bioBDS Science Fund of the Republic of Serbia (No. 7750294).

References

1. Makarova K.S. et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*. 2020;18:67-83
2. Sayers E.W. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2022;50:D20-D26
3. Mitrofanov A. et al. CRISPRidentify: identification of CRISPR arrays using machine learning approach. *Nucleic Acids Res*. 2021;49:e20
4. Padilha V.A. et al. CRISPRcasIdentifier: Machine learning for accurate identification and classification of CRISPR-Cas systems. *GigaScience*. 2020;9:giaa062
5. Virus-Host D.B., Virus-Host D.B. Website, 2020. https://www.genome.jp/virushostdb
6. Wang M. et al. VRprofile2: detection of antibiotic resistance-associated mobilome in bacterial pathogens. *Nucleic Acids Res*. 2022;50:W768-W73

doi 10.18699/bgrs2024-8.4-21

Analysis of the amino acid sequences aminoacyl-phosphatidylglycerol synthases of gram-positive bacteria

Yaburova E.V.*, Polyudova T.V.
Perm Federal Research Center of the Ural Branch of the Russian Academy of Sciences (PFRC UB RAS), Perm, Russia
** yaburova_k@mail.ru*

Key words: aminoacylated phospholipids; membrane proteins; antibacterial resistance; multiple peptide resistance factor (MprF)

Motivation and Aim: Aminoacyl-phosphatidylglycerol synthases (aaPGSs) serve as defense mechanism in bacterial cells. These enzymes catalyze the transfer of amino acids from aminoacyl-tRNAs to phosphatidylglycerol (PG) and transport the resulting product into the outer layer of the bacterial membrane. Through the aminoacylation of PG, bacteria are able to alter the overall negative charge of their cell envelope, decreasing its affinity for charged substances and adapting to different environmental conditions. This process is crucial for the pathogenicity of certain microbes, providing them with protection against antimicrobial peptides from host organisms and therapeutic antibiotics [1].

An integral membrane protein aaPGSs known as Multiple Peptide Resistance Factor (MprF) exists in the form of homodi- and tetramers. Changes in the structure of this protein can sometimes restore sensitivity to antibacterial compounds in resistant microorganisms [2]. Assessing the prevalence of this defense mechanism among the two largest bacterial phyla with a gram-positive phenotype, *Bacillota* and *Actinomycetota* will enhance our understanding of its role in adaptive processes and help identify strategies to overcome it [3].

Methods and Algorithms: The bioinformatic analysis was performed by the method of multiple alignment of amino acid sequences of the MprF protein using the BLAST algorithm blastp (the database RefSeq Select proteins). The search for proteins was carried out among the reference sequences of the phyla *Bacillota* and *Actinomycetota* bacteria available in the NCBI database (https://www.ncbi.nlm.nih.gov). Sequences were retrieved by BLAST [4] using the sequence Lysyl-PGS from *Bacillus subtilis* (YP_003097695.1) as a template for *Bacillota* and the LysylX-PGS *Mycobacterium tuberculosis* (NP_216156.1) as a template for *Actinomycetota*. AaPGSs sequences displaying an E value < 10E-6.

Results: To date, putative aaPGS shave been identified in 921 distinct species of the 129 genera of the *Bacillota* and 1,453 distinct species of the 208 genera of the *Actinomycetota*. Approximately 250 genera are considered to be within the *Bacillota* phylum and more 400 genera – the *Actinomycetota* phylum. However, the protein reference database continues to grow, and a gap in the number of species does not necessarily mean that they lack aaPGSs.

AaPGSs of the *Bacillota* consist of two separate domains to synthesize and translocate aminoacyl phospholipids to the outer leaflets of bacterial membranes. Both domains vary in size and in sequence. AaPGSs found in *Actinomycetota* (*Actinomycetales*, *Micrococcales*, *Mycobacteriales* and others), are fused with an additional Lys-RS

domain that exhibits all of the features associated with tRNA lysylation such as a tRNA anticodon binding domain along with all the conserved positions in Lys-RSs necessary for substrate binding (Lys, ATP, and tRNA) [2]. However, the analysis of the peptide sequences extracted from the database showed there are aaPGSs of *Bifidobacteriales* and *Kineosporiales* (both orders belong to class *Actinomycetia*, phylum *Actinomycetota*), which is characterized by the absence of an additional Lys-RS domain.

Conclusion: The study of aminoacylphosphatidylglycerol synthases (aaPGSs) and their role as defense mechanisms in bacterial cells, sheds light on the adaptive processes of bacteria in response to environmental challenges. The bioinformatic analysis conducted to assess the prevalence of aaPGSs among the *Bacillota* and *Actinomycetota* phyla has revealed significant insights into the structural variations and functional diversity of these enzymes. The identification of aaPGSs in a wide range of bacterial species underscores the importance of these enzymes in providing protection against antimicrobial agents. Further research in this area will not only enhance our understanding of bacterial adaptation but also contribute to the development of strategies to overcome bacterial resistance mechanisms.

References

1. Roy H., Ibba M. Broad range amino acid specificity of RNA-dependent lipid remodeling by multiple peptide resistance factors. *J Biol Chem.* 2009;284(43):29677-29683. doi 10.1074/jbc.M109.046367
2. Song D., Jiao H., Liu Z. Phospholipid translocation captured in a bifunctional membrane protein MprF. *Nat Commun.* 2021;12(1):2927. doi 10.1038/s41467-021-23248-z
3. Oren A., Garrity G.M. Valid publication of the names of forty-two phyla of prokaryotes. *Int J Syst Evol Microbiol.* 2021;71(10). doi 10.1099/ijsem.0.005056
4. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389-3402. doi 10.1093/nar/25.17.3389

9 Симпозиум «Биомедицина, биоинформатика и системная компьютерная биология»

Symposium “Biomedicine, bioinformatics and systems computational biology”

doi 10.18699/bgrs2024-9.1-01

Биоинформационный анализ генов миопии

Алейникова А.В.[1]*, Савина Е.А.[1], Орлов Ю.Л.[1, 2]

[1] *Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

**aleyn.a@mail.ru, orlov@bionet.nsc.ru*

Ключевые слова: медицинская информатика, медицинская генетика, заболевания глаз, миопия, базы данных, генные онтологии

Мотивация и цель: Миопия является мультифакториальным заболеванием, это значит, что развитие данной болезни зависит и от генетических особенностей индивида, и от образа его жизни [1]. Таким образом, сложное и неразрывное взаимодействие генетических факторов с окружающей средой приводит к началу близорукости и ее дальнейшей прогрессии. Для наибольшего понимания вопроса миопии, рассмотрим подробнее генетическую основу близорукости.

Методы и алгоритмы: Обзор литературы выполнен с помощью PubMed, PMC, базы elibrary.ru. Для определения генов, ассоциированных с миопией, использовались базы данных генетической информации OMIM, NCBI GenBank, GeneOntology.org и публикации GWAS, опубликованные популяционные данные. Списки генов анализировались с помощью ресурсов GeneCards.org и STRING-DB.org, metascape.org.

Результаты: Собрана научная литература по генетике миопии. Миопия может передаваться по аутосомно-доминантному, по аутосомно-рецессивному и по Х-сцепленному рецессивному типу наследования [2–4]. При наследовании по аутосомно-доминантному типу болезнь не проявляется долгое время. Обычно заболевание начинает беспокоить пациентов в подростковом, либо школьном возрасте. В таком случае клиническое течение болезни более мягкое. При наследовании по аутосомно-рецессивному типу болезнь сразу проявляет себя в раннем детском возрасте. Клиническая картина сопровождается множеством сопутствующих осложнений. Однако учеными было доказано, что при передаче по аутосомно-доминантному типу наследования неблагоприятные факторы внешней среды имеют гораздо большую роль в формировании патологий, чем при передаче аутосомно-рецессивной формы близорукости. При Х-сцепленном рецессивном наследовании заболевание проявляется у лиц мужского пола. Лица женского пола являются носителями мутаций, болезнь у них не находит фенотипических проявлений.

В целом близорукость представляет собой сложное заболевание, которое контролируется множеством взаимосвязанных генов. Наиболее распространенные формы легкой и средней миопии передаются как сложные многогенные признаки, тяжелые же формы миопии передаются как моногенные дефекты и чаще всего сопровождаются сопутствующими осложнениями и аномалиями. Именно поэтому в основе изучения генетики миопии существуют значительные сложности: множество генов, отвечающих за развитие миопии характеризуются низкой частотой встречаемости или слабо выраженным эффектом проявления, что вызывает затруднения в их определении с помощью

классических генетических методов. По данным ученых, исследования полногеномного поиска ассоциаций с применением методики GWAS выявило множество локусов, которые содержат гены, влияющие на течение и развитие миопии. Однако, по оценкам исследователей, на эти варианты локусов приходится менее 5 % от всех популяционных изменений. Данное явление называется проблема отсутствующей наследственности. Причиной этого может являться ограниченная статистическая мощность и наличие большого количества редких вариантов с низкой частотой встречаемости, которые не были идентифицированы методом GWAS. Известно, что статистическая мощность метода GWAS экспоненциально снижается с уменьшением частоты минорных аллелей, из-за этого традиционный полногеномный поиск ассоциаций GWAS не очень подходит для идентификации редких вариантов и выявляет в основном общие варианты.
На сегодняшний день идентифицированы 26 локусов хромосом, которые ассоциированы с проявлением миопии. В частности, найдена ассоциация для генов COL2A1, COL1A1, COL18A1, FBN1, NYX, CACNA1F, GRM6, LRIT3, RP2, RPGR, ADAMTS10, ADAMTS17, ADAMTS18, VPS13B, OPN1LW, SCO2, ZNF644, CCDC111, LRPAP1, SLC39A5, P4HA2, ARR3, CPSF1, BSG, CTSH, DZIP1, LEPREL1, LOXL3, NDUFAF7, TNFRSF21, XYLT1. Рассмотрены категории генных онтологий и функциональная аннотация найденных генов.
Выводы: Подводя итоги, миопия представляет собой нарушение рефракции, которое происходит чаще всего из-за увеличения осевой длины глаза. Это мультифакториальное заболевание, большую роль в котором играет влияние факторов среды. Близорукость носит эпидемиологический характер и широко распространена по всему миру, особенно в странах Восточной Азии. На данный момент еще не создано действенного решения в излечении миопии, в связи со сложностью сбора информации по всем генетическим аспектам. Именно поэтому стоит продолжать изучать генетическую основу данной патологии для нахождения соответствующих потенциальных терапевтических генов-мишеней.
Финансирование: Исследование поддержано РНФ (грант 24-24-00563, «Разработка цифровых образовательных программ в биомедицине»).

Bioinformatics analysis of myopia genes

Aleynikova A.V.[1]*, Savina E.A.[1*], Orlov Y.L.[1, 2]

[1] *Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *aleyn.a@mail.ru, orlov@bionet.nsc.ru*

Key words: medical informatics, medical genetics, eye diseases, myopia, databases, gene ontologies

Motivation and Aim: Myopia is a multifactorial disease, which means that the development of this disease depends on both the genetic characteristics of the individual and his lifestyle [1]. Thus, the complex and inextricable interaction of genetic factors with the environment leads to the onset of myopia and its further progression. For the best understanding of the issue of myopia, let's take a closer look at the genetic basis of myopia.

Methods and Algorithms: The literature review was carried out using PubMed, PMC, database elibrary.ru. OMIM and NCBI GenBank genetic information databases were used to identify genes associated with myopia, GeneOntology.org and GWAS publications, published population data. Gene lists were analyzed using resources GeneCards.org, STRING-DB.org, and metascape.org.
Results: Scientific literature on the genetics of myopia has been collected. Myopia can be transmitted by autosomal dominant, autosomal recessive, and X-linked recessive inheritance [2–4]. When inherited by an autosomal dominant type, the disease does not manifest itself for a long time. Usually, the disease begins to bother patients in adolescence or school age. In this case, the clinical course of the disease is milder. When inherited by an autosomal recessive type, the disease immediately manifests itself in early childhood. The clinical picture is accompanied by many concomitant complications. However, scientists have proved that in the transmission of an autosomal dominant type of inheritance, adverse environmental factors play a much greater role in the formation of pathologies than in the transmission of an autosomal recessive form of myopia. With X-linked recessive inheritance, the disease manifests itself in males. Female individuals are carriers of mutations, the disease does not find phenotypic manifestations in them.
In general, myopia is a complex disease that is controlled by many interconnected genes. The most common forms of mild and moderate myopia are transmitted as complex multigenic signs, while severe forms of myopia are transmitted as monogenic defects and are most often accompanied by concomitant complications and abnormalities. That is why there are significant difficulties in studying the genetics of myopia: many genes responsible for the development of myopia are characterized by a low frequency of occurrence or a weakly pronounced manifestation effect, which causes difficulties in determining them using classical genetic methods. According to scientists, studies of genome-wide association search using the GWAS technique have revealed many loci that contain genes that affect the course and development of myopia. However, according to the researchers, these locus variants account for less than 5% of all population changes. This phenomenon is called the problem of missing heredity. The reason for this may be limited statistical power and the presence of a large number of rare variants with a low frequency of occurrence that have not been identified by the GWAS method. It is known that the statistical power of the GWAS method decreases exponentially with a decrease in the frequency of minor alleles, because of this, the traditional genome-wide search for GWAS associations is not very suitable for identifying rare variants and reveals mainly common variants.
To date, scientists have identified 26 chromosome loci that are associated with the manifestation of myopia. In particular, an association was found for the genes COL2A1, COL1A1, COL18A1, FBN1, NYX, CACNA1F, GRM6, LRIT3, RP2, RPGR, ADAMTS10, ADAMTS17, ADAMTS18, VPS13B, OPN1LW, SCO2, ZNF644, CCDC111, LRPAP1, SLC39A5, P4HA2, ARR3, CPSF1, BSG, CTSH, DZIP1, LEPREL1, LOXL3, NDUFAF7, TNFRSF21, XYLT1. The categories of gene ontologies and the functional annotation of the found genes were considered.
Conclusion: To summarize, myopia is a refractive error that occurs most often due to an increase in the axial length of the eye. This is a multifactorial disease, in which the influence of environmental factors plays an important role. Myopia is epidemiological in nature and is widespread around the world, especially in East Asian countries. At the moment, an effective solution has not yet been created in the treatment of myopia, due

to the complexity of collecting information on all genetic aspects. That is why it is worth continuing to study the genetic basis of this pathology in order to find appropriate potential therapeutic target genes.

Funding: The study was supported by the Russian Science Foundation (grant 24-24-00563).

Список литературы/References

1. Куликов А.Н., Чурашов С.В., Рейтузов В.А. Молекулярно-генетические аспекты патогенеза прогрессирующей миопии. *Офтальмологические ведомости.* 2018;11(3):48-56. doi 10.17816/OV11348-56
   [Kulikov A.N., Churashov S.V., Reituzov V.A. Molecular genetic aspects of complicated myopia pathogenesis. *Ophthalmology Journal*. 2018;11(3):48-56. doi 10.17816/OV11348-56 (in Russian)]
2. Haarman A.E.G., Thiadens A.A.H.J., van Tienhoven M. et al. Whole exome sequencing of known eye genes reveals genetic causes for high myopia. *Hum Mol Genet.* 2022;31(19):3290-3298. doi 10.1093/hmg/ddac113
3. Chen C., An G., Yu X. et al. Screening Mutations of the Monogenic Syndromic High Myopia by Whole Exome Sequencing From MAGIC Project. *Invest Ophthalmol Vis Sci.* 2024;65(2):9. doi 10.1167/iovs.65.2.9
4. Cai X.B., Shen S.R., Chen D.F., Zhang Q., Jin Z.B. An overview of myopia genetics. *Exp Eye Res.* 2019;188:107778. doi 10.1016/j.exer.2019.107778

doi 10.18699/bgrs2024-9.1-02

Применение методов интерпретируемого машинного обучения для поиска транскриптомных биомаркеров болезни Альцгеймера

Асаинов Д.Т.[1]*, Кириллов Б.А.[2, 3]

[1] *Кафедра биохимии, биотехнологии и фармакологии, Институт фундаментальной медицины и биологии, Казанский (Приволжский) федеральный университет, Казань, Россия*

[2] *Центр технологий материалов, Сколковский институт науки и технологий, Москва, Россия*

[3] *Центр высокоточного редактирования и генетических технологий для биомедицины, Институт биологии гена Российской академии наук, Москва, Россия*

** dasainov@inbox.ru*

Ключевые слова: болезнь Альцгеймера; машинное обучение; биомаркеры; транскриптом

Мотивация и цель: Диагностика болезни Альцгеймера (БА) представляет собой продолжительный процесс [1], включающий в себя изучение анамнеза, тесты, направленные на оценку психического статуса пациента, инвазивные анализы, необходимые для оценки уровня бета-амилоида в спинномозговой жидкости, и методы нейровизуализации. При этом ранняя диагностика с применением классических методов затруднена из-за неоднородности симптомов и их неоднозначного распределения по времени появления, что наряду с необходимостью дифференцирования БА от других видов деменции и невозможностью однозначно подтвердить диагноз до гистологического исследования мозга говорит о серьезном потенциале альтернативных методов диагностики БА. К ним относятся и модели машинного обучения, которые открывают возможности как для высокоточной ранней диагностики, так и для изучения аспектов патогенеза. Однако существует характерная для многих моделей проблема «черного ящика» – иначе говоря, отсутствия информации о том, как модели машинного обучения принимают решения, что ограничивает их применение в клинической практике. Для решения этой проблемы были созданы методы Explainable Machine Learning (xML, IML [2]) – объяснимого или интерпретируемого машинного обучения. Методы xML позволяют получить информацию о механизмах принятия решений и вкладах в прогнозы отдельных входных признаков, что должно повысить уровень доверия к новым для медицинского сообщества способам диагностики и облегчить получение информации о потенциальных драйверах БА.

Методы и алгоритмы: Для обучения моделей мы использовали фреймворк Ludwig [3], ориентированный на методы глубокого обучения. Используемый датасет представляет собой полученные с чипа Illumina Human HT-12 v4 данные об экспрессии РНК в средней височной извилине у пациентов с диагностированной БА и контрольной группы [4]. Для генерации объяснений предсказаний мы использовали алгоритм Accumulated Local Effects (ALE) в составе библиотеки Alibi [5].

Результаты: Мы обучили ряд моделей, прогнозирующих наличие БА, лучшие значения метрик которых составляют: сбалансированная точность – 0.924,

специфичность – 0.886, коэффициент корреляции Мэтьюза – 0.85. Объяснения, полученные с помощью ALE, позволили выделить ряд транскриптов, являющихся особо яркими входными признаками для обученных моделей. Соответствующие гены, представляющие интерес, вовлечены в процессы свертывания крови, возникновения воспалительных реакций, биосинтеза дофамина и поддержания функции макрофагов. В ходе анализа литературы были найдены подтверждения связи этих генов с болезнью Альцгеймера.
Выводы: Результаты нашего исследования демонстрируют возможность извлечения из высокоточных моделей машинного обучения информации о потенциально задействованных в патогенезе БА генах, что потенциально может повысить уровень доверия к подобным моделям в клинической практике и оптимизировать процедуры подтверждения диагноза и ранней диагностики.

Application of Explainable Machine Learning methods to find transcriptomic biomarkers of Alzheimer's disease

Asainov D.[1]*, Kirillov B.[2, 3]

[1] *Department of Biochemistry, Biotechnology and Pharmacology, Institute of Fundamental Medicine and Biology, Kazan (Volga Region) Federal University, Kazan, Russia*

[2] *Materials Technology Center, Skolkovo Institute of Science and Technology, Moscow, Russia*

[3] *Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia*

* *dasainov@inbox.ru*

Key words: Alzheimer's disease; machine learning; biomarkers; transcriptome

Motivation and Aim: Diagnosis of Alzheimer's disease (AD) is a lengthy process [1], that includes gathering of anamnesis, tests for the patient's mental status, invasive tests necessary to assess the level of beta-amyloid in the cerebrospinal fluid, and neuroimaging methods. However, early diagnosis using classical methods is complicated by the heterogeneity of symptoms in a cohort of patients with Alzheimer's disease and their ambiguous distribution by time of onset, which, along with the need to differentiate AD from other types of dementia and the inability to confirm the diagnosis before histological examination of the brain, indicates the serious potential of alternative methods of diagnosing AD. Such methods include machine learning models, which offer opportunities both for high-precision early diagnosis and for studying the aspects of pathogenesis. However, the majority of models possess a "black box" problem – the lack of information about how machine learning models make decisions. Such lack of information limits the applicability of machine learning models in clinical practice. Explainable Machine Learning (xML, IML [2]) methods – explainable or interpretable machine learning – have been created to address this problem. xML methods provide information about decision-making mechanisms and contributions of individual input features to the predictions, which should increase the level of confidence in new diagnostic methods for the medical community and facilitate the acquisition of information about potential drivers of AD.
Methods and Algorithms: To train the models, we use the Ludwig framework [3] focused on deep learning methods. The dataset we use is RNA expression data for the middle temporal gyrus of patients with diagnosed AD and controls [4] obtained from the

Illumina Human HT-12 v4 chip. We use the Accumulated Local Effects (ALE) algorithm as part of the Alibi library [5] to generate explanations for our predictions.
Results: We have trained a number of models predicting the presence of AD, the best metrics of which are: Balanced Accuracy – 0.924, Precision – 0.886, Matthews Correlation Coefficient – 0.85. The explanations obtained with ALE allowed us to identify several transcripts that are particularly important input features for the trained models. Relevant genes of interest, found by our ALE test, are involved in blood coagulation, the occurrence of inflammatory responses, dopamine biosynthesis and maintenance of macrophage function. We found literature that supports the association of these genes with Alzheimer's disease.
Conclusion: The results of our study demonstrate the feasibility of extracting information about genes potentially involved in AD pathogenesis from machine learning models, potentially increasing the level of confidence in such models in clinical practice and optimizing diagnosis confirmation and early diagnosis procedures.

Список литературы/References

1. McKhann G.M., Knopman D.S., Chertkow H. et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7(3):263-269
2. Belle V., Papantonis I. Principles and practice of explainable machine learning. *Front Big Data*. 2021;4:688969. doi 10.3389/fdata.2021.688969
3. Molino P., Dudin Y., Miryala S.S. Ludwig: a type-based declarative deep learning toolbox. *arXiv*. 2019. doi 10.48550/arXiv.1909.07930
4. Piras I.S., Krate J., Delvaux E., Nolz J. et al. Transcriptome changes in the Alzheimer's disease middle temporal gyrus: importance of RNA metabolism and mitochondria-associated membrane genes. *J Alzheimers Dis.* 2019;70(3):691-713
5. Klaise J., Looveren A.V., Vacanti G., Coca A. Alibi explain: algorithms for explaining machine learning models. *J Mach Learn Res.* 2021;22(1):8194-8200

doi 10.18699/bgrs2024-9.1-03

Конкурирующие эндогенные РНК как маркеры для немелкоклеточного рака легких

Бурденный А.М.[1, 2], Логинов В.И.[1, 3], Пронина И.В.[1], Брага Э.А.[1, 3]

[1] *Научно-исследовательский институт общей патологии и патофизиологии, Москва, Россия*

[2] *Институт биохимической физики им. Н.М. Эмануэля РАН, Москва, Россия*

[3] *Медико-генетический научный центр им. академика Н.П. Бочкова, Москва, Россия*

* *burdennyy@gmail.com*

Ключевые слова: аденокарцинома, плоскоклеточный рак, длинная некодирующая РНК, опухолевые маркеры

Мотивация и цель. Рак легких является наиболее часто диагностируемым злокачественным новообразованием и ведущим источником неблагоприятного исхода от рака [1]. Этот вид злокачественных новообразований, также как и другие, имеет гетерогенную природу и представлен двумя большими классами: мелкоклеточный рак легкого (МРЛ) и немелкоклеточный рак легкого (НМРЛ). На НМРЛ приходится не менее 85 % всех случаев рака легких. Этот вид рака подразделяется на аденокарциному, плоскоклеточный рак и крупноклеточный рак [2]. Отметим, что первые два класса имеют свои существенные особенности в диагностике, патогенезе, локализации, визуализации, метастазировании и лечении [3]. В связи с этим, актуальной задачей является поиск возможных молекулярных маркеров, которые могут помочь однозначно определить, как класс, так и подтип рака лёгких. Таким инструментом могут стать маркеры эпигенетической регуляции, представленные некодирующими РНК. Этот класс молекул характеризуется тем, что вовлечен в управление всех клеточных процессов, и нарушения в работе этих молекул могут стать ключевым фактором развития как рака в целом, так и НМРЛ в частности. Согласно теории конкурирующих эндогенных РНК, длинные некодирующие РНК (днРНК) конкурируют с мРНК белок-кодирующих генов за связывание с микроРНК (миРНК), тем самым влияя на уровень мРНК [4].

Целью настоящего исследования являлся поиск новых маркеров немелкоклеточного рака легкого на основе конкурентных эндогенных РНК.

Методы и алгоритмы. Данные об экспрессии генов у пациентов с НМРЛ получены из Атласа генома рака (TCGA, https://portal.gdc.cancer.gov/). Были отобраны дифференциально экспрессируемые (ДЭ) гены информационных (матричных) РНК (мРНК), микроРНК (миРНК) и длинных некодирующих РНК (днРНК), удовлетворяющие правилу |Log2 FoldChange|>3 и $p < 0.01$. Связь днРНК с раком была изучена по базам LncRNADisease 2.0 (www.cuilab.cn/lncrnadisease) и lncATLAS (https://lncatlas.crg.eu/). Транскриптомный ландшафт днРНК был идентифицирован с помощью Lnc2Cancer 2.0 (www.bio-bigdata.net/lnc2cancer).

Для изучения биологических функций ДЭмРНК были использованы системы «Генная онтология» (Gene ontology, https://geneontology.org/) и энциклопедия генов и геномов (Киото) (KEGG, https://www.genome.jp/kegg/).

В настоящем исследовании была построена сеть (днРНК-миРНК-мРНК) конкурирующих эндогенных РНК, которая основана на гипотезе о том, что днРНК являются «губкой» для миРНК, тем самым осуществляя регуляцию активности

мРНК белок-кодирующих генов. Для получения данных о возможных взаимодействиях миРНК–мРНК и миРНК–днРНК в программу starBase2.0 (http://starbase.sysu.edu.cn) были введены данные о днРНК, миРНК, и мРНК с аномальным уровнем экспрессии (кратностью изменений >3.0, <0.33 и $p < 0.01$). Для прогнозирования взаимодействий днРНК-миРНК использовались инструменты miRanda (http://www.microrna.gr/) и miRCode (http://www.mircode.org/), а миРНК-мРНК – miRWalk 2.0 (http://mirwalk.umm.uni-heidelberg.de/), TargetScan (http://www.targetscan.org/) и miRTarBase (https://mirtarbase.cuhk.edu.cn). Для построения сети конкурирующих эндогенных РНК были выбраны миРНК, которые имеют негативную корреляцию с экспрессией днРНК и мРНК. Для конструирования и визуализации сети днРНК-миРНК-мРНК использовали инструментальную среду Cytoscape v3.0 (http://www.cytoscape.org/).

Результаты. При сравнении опухолевых тканей и прилегающих гистологически неизмененных тканей пациентов с раком легких из базы данных TCGA мы обнаружили в общей сложности 735 ДЭмРНК, 12 ДЭмиРНК и 107 ДЭднРНК. Дополнительным этапом отбора было выявление в области не далее 2000 п.н. от 5’ конца гена или старта транскрипции обособленных CpG островков с использованием базы Epigenomics https://www.ncbi.nlm.nih.gov/gene/. В результате из вышеперечисленного осталось около 60 днРНК, 6 миРНК и около 250 мРНК белок-кодирующих генов.

В результате функционального анализа было показано, что мРНК с повышенным уровнем экспрессии в основном участвуют в митотическом клеточном цикле и делении клеток, а мРНК с пониженной экспрессией в основном участвуют в передаче сигнала в клетке, клеточной адгезии и иммунном ответе. Мы предположили, что гены участвующие в клеточном цикле, сигнальных путях p53, PI3K-Akt, Ca^{2+} могут быть вовлечены в прогрессию НМРЛ.

Всего в сеть конкурентных эндогенных РНК были вовлечены 6 днРНК (*ADAMTS9-AS2, HAND2-AS1, HOTAIRM1, MEG3, SNHG12, NCK1-AS1*) 4 миРНК (miR-124-3p, miR-203a-3p, miR-125b-5p, miR-129-5p) и 26 мРНК. В опухолевых тканях уровень *NCK1-AS1* был повышен, тогда как уровень *ADAMTS9-AS2, HAND2-AS1, HOTAIRM1, MEG3, SNHG12* значимо снижался в сравнении с таковым прилежащих гистологически неизмененных тканях. Количество узлов и линий в сети составило 36 и 60 соответственно. Из 6 днРНК количество миРНК, связанных с *ADAMTS9-AS2, HAND2-AS1, MEG3* было больше, чем у других днРНК, что указывает на их важную роль прогрессии НМРЛ. Из них 3 днРНК, 2 микроРНК и 15 мРНК были тесно связаны с общей выживаемостью.

Это исследование имеет некоторые ограничения. Во-первых, мы не проверяли эти новые биомаркеры с использованием дополнительных выборок данных. Во-вторых, мы не проводили молекулярно-биологические эксперименты для дальнейшей проверки функций и механизмов регуляции идентифицированных эндогенных РНК у пациентов с НМРЛ. Необходимы дальнейшие исследования с использованием клинических образцов и функциональных экспериментов на клеточных линиях для подтверждения наших выводов.

Выводы. Таким образом, мы построили сеть днРНК-миРНК-мРНК у пациентов с НМРЛ и определили 15 мРНК, 2 микроРНК и 3 днРНК в качестве факторов прогноза общей выживаемости в вышеупомянутой популяции. Настоящее исследование позволяет глубже оценить роли каждого элемента сети эндогенных

РНК, у пациентов с НМРЛ, а некоторые из генов этой сети могут быть использованы в качестве новых потенциальных терапевтических мишеней.
Финансирование: Исследование поддержано государственным заданием FGFU-2023-0001.

Competing endogenous RNAs as non-small cell lung cancer markers

Burdennyy A.M.[1, 2]*, Loginov V.I.[1, 3], Pronina I.V.[1], Braga E.A.[1, 3]

[1] *Institute of General Pathology and Pathophysiology, Moscow, Russia*

[2] *Emanuel Institute for Biochemical Physics, Moscow, Russia*

[3] *Bochkov Research Centre for Medical Genetics, Moscow, Russia*

* *burdennyy@gmail.com*

Key words: lung adenocarcinoma, squamous cell lung carcinoma, long non-coding RNA, oncomarkers

Motivation and Aim. Lung cancer is the most frequently diagnosed malignant neoplasm and the leading source of cancer adverse outcome [1]. This type of malignant neoplasm, as well as others, has a heterogeneous nature and is represented by two large classes: small cell lung cancer (MRL) and non-small cell lung cancer (NSCLC). NSCLC accounts for at least 85% of all lung cancer cases. This type of cancer is divided into adenocarcinoma, squamous cell carcinoma and large cell carcinoma [2]. It should be noted that the first two classes have their own significant features in diagnosis, pathogenesis, localization, visualization, metastasis and treatment [3]. In this regard, an urgent task is to search for possible molecular markers that can help uniquely identify both the class and subtype of lung cancer. Markers of epigenetic regulation, represented by non-coding RNAs, can become such a tool. This class of molecules is characterized by the fact that it is involved in the regulation of all cellular processes, and dysregulation in the work of these molecules can become a key factor in the development of both cancer in general and NSCLC in particular. According to the theory of competing endogenous RNAs, long non-coding RNAs (lncRNAs) compete with the mRNAs of protein-coding genes for binding to microRNAs (miRNAs), thereby affecting the mRNA level [4].

The aim of this study was to search for new markers of non-small cell lung cancer based on competitive endogenous RNAs.

Methods and Algorithms. Gene expression dataset of patients with NSCLC were obtained from the Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/). Differentially expressed (DE) genes of informational (matrix) RNAs (mRNAs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) that satisfying the rule $|Log2FoldChange|>3$ and $p<0.01$ were selected. The association of lncRNAs with cancer was studied using LncRNADisease 2.0 (www.cuilab.cn/lncrnadisease) and lncATLAS (https://lncatlas.crg.eu/) databases. The transcriptomic landscape of lncRNAs was identified using Lnc2Cancer 2.0 (www.bio-bigdata.net/lnc2cancer).

Gene ontology (https://geneontology.org/) and the Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg/) systems were used to study the biological functions of DEmRNAs.

In present study, a network (lncRNA-miRNA-mRNA) of competing endogenous RNAs was constructed, which is based on the hypothesis of lncRNAs were to be the "sponge" for miRNAs, thereby regulating the activity of mRNA protein-coding genes. To obtain data of possible miRNA–mRNA and miRNA–lncRNA interactions, the data of abnormal lncRNAs, miRNAs, and mRNAs expression levels (multiplicity of changes >3.0, <0.33, and $p<0.01$) were entered in the starBase2.0 program (http://starbase.sysu.edu.cn). Database miRanda (http://www.microrna.gr/) and miRCode (http://www.mircode.org/) tools were used to predict lncRNA-miRNA interactions, and miRWalk 2.0 (http://mirwalk.umm.uni-heidelberg.de/), TargetScan (http://www.targetscan.org/) and miRTarBase (https://mirtarbase.cuhk.edu.cn) databases were used for miRNA-mRNA interactions. To build a network of competing endogenous RNAs, miRNAs that have a negative correlation with the expression of lncRNA and mRNA were selected. Cytoscape v3.0 instrumental environment was used to construct and visualize the lncRNA-miRNA-mRNA network (http://www.cytoscape.org/)/.

Results. When comparing tumor tissues and adjacent histologically unchanged tissues of lung cancer patients from the TCGA database, we found a total of 735 DEmRNAs, 12 DEmiRNAs and 107 DElncRNAs. An additional stage of selection was the identification of isolated CpG islands in the region no farther than 2000 bp from the 5' end or the start of transcription of the gene using the Epigenomics database https://www.ncbi.nlm.nih.gov/gene/. As a result, about 60 lncRNAs, 6 miRNAs and about 250 mRNAs of protein-coding genes remained from the above mentioned.

As a result of functional analysis, it was shown that mRNAs with an increased level of expression are mainly involved in the mitotic cell cycle and cell division, and mRNAs with reduced expression are mainly involved in signal transmission in the cell, cell adhesion and immune response. We hypothesized that genes involved in the cell cycle, signaling pathways p53, PI3K-Akt, Ca2+ may be involved in the progression of NSCLC. A total of 6 lncRNAs (ADAMTS9-AS2, HAND2-AS1, HOTAIRM1, MEG3, SNHG12, NCK1-AS1), 4 miRNAs (miR-124-3p, miR-203a-3p, miR-125b-5p, miR-129-5p) and 26 mRNAs were involved in the network of competitive endogenous RNAs. In tumor tissues, the level of NCK1-AS1 was increased, whereas the level of *ADAMTS9-AS2, HAND2-AS1, HOTAIRM1, MEG3, SNHG12* significantly decreased in comparison with that of adjacent histologically unchanged tissues. The number of nodes and lines in the network was 36 and 60, respectively. From the 6 lncRNAs, the number of miRNAs associated with *ADAMTS9-AS2, HAND2-AS1*, and *MEG3* was higher than that for other lncRNAs, indicating their important role in NSCLC progression. Of these set, 3 lncRNAs, 2 microRNAs, and 15 mRNAs were closely associated with overall survival. This study has some limitations. Firstly, we did not test these new biomarkers using additional data samples. Secondly, we did not conduct molecular biological experiments to further verify the functions and mechanisms of regulation of identified endogenous RNAs in patients with NSCLC. Further studies using clinical samples and functional experiments on cell lines are needed to confirm our findings.

Conclusion. Thus, we constructed a lncRNA-miRNA-mRNA network in patients with NSCLC and identified 15 mRNAs, 2 microRNAs and 3 lncRNAs as predictors of overall survival in the above-mentioned population. This study allows us to more deeply evaluate the roles of each element of the endogenous RNA network in patients with NSCLC, and some of the genes in this network may be used as new potential therapeutic targets.

Funding: The study is supported by the FGFU-2023-0001 state assignment.

Список литературы/References

1 Bray F., Ferlay J., Soerjomataram I., Siegel R.L., Torre L.A., Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424

2 Inamura K. Lung cancer: understanding its molecular pathology and the 2015 WHO classification. *Front Oncol*. 2017;7:193

3 Chen J.W., Dhahbi J. Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. *Sci Rep*. 2021;11(1):13323. doi 10.1038/s41598-021-92725-8

4 Yu Y., Ren K. Five long non-coding RNAs establish a prognostic nomogram and construct a competing endogenous RNA network in the progression of non-small cell lung cancer. *BMC Cancer*. 2021;21(1):457. doi 10.1186/s12885-021-08207-7

doi 10.18699/bgrs2024-9.1-05

Сплайсинговые изоформы *SIRT1* в печени у пациентов с метаболическим синдромом – ожидание/реальность

Горбачёва А.М.*, Воронова С.С., Боград М.М., Вульф М.А., Литвинова Л.С.
Балтийский федеральный университет имени И. Канта, Калининград, Россия
** anmigorbacheva@gmail.com*

Ключевые слова: SIRT1, изоформы SIRT1, метаболический синдром

Мотивация и цель: Сиртуин 1 (SIRT1) – НАД+/НАДФН-зависимая деацетилаза, участвующая в модуляции липогенеза и липолиза, биогенеза митохондрий через регуляцию активности транскрипционных факторов PGC-1α, PPAR-α и PPAR-γ. SIRT1 многократно рассматривался в контексте метаболического синдрома (МС), но его изоформы практически не изучены [1]. Изоформы SIRT1 образуются в результате альтернативного сплайсинга: изоформа 1 (V1) имеет два ядерных сигнала локализации, две другие изоформы (V2 и V3) располагаются в цитоплазме. Таким образом, *целью* данной работы явилось биоинформатическое предсказание взаимодействий изоформ SIRT1 с транскрипционными факторами PGC-1α, PPAR-α, PPAR-γ, изучение уровня экспрессии изоформ *SIRT1* и перечисленных транскрипционных факторов в печени у пациентов с МС.
Методы и алгоритмы: Белок-белковые взаимодействия были предсказаны с помощью инструмента PEPPI [2]. Использовались аминокислотные последовательности из UniProt: SIRT1 V1 (Q96EB6), SIRT1 V2 (E9PC49), SIRT1 V3 (B0QZ35), PGC-1α (Q9UBK2), PPAR-α (Q07869), PPAR-γ (P37231) [3].
Были исследованы биоптаты печени от 63 доноров (средний возраст = 45 ± ± 8.48 года, 40 женщин и 23 мужчины), которые были разделены на две группы: 1) контрольная группа (среднее значение ИМТ = 19.8 ± 1.4 кг/м^2, 6 женщин, 8 мужчин); 2) пациенты с МС (среднее значение ИМТ = 44.3 ± 12.3 кг/м^2, 34 женщины, 15 мужчин), критерии включения: минимум два компонента МС в анамнезе и кроме того, тощаковая глюкоза в крови >5.5 ммоль/л и индекс атерогенности >3. Уровень экспрессии генов интереса изучали с помощью количественной ОТ-ПЦР с использованием SYBR Green (*Evrogen*, Россия).
Результаты: Было предсказано, что все 3 изоформы SIRT1 связываются с транскрипционным фактором PPAR-α. В то же время V2 и V3 связываются с PGC-1α, PPAR-γ (табл. 1).

Таблица 1. Биоинформатический и экспериментальный анализ взаимодействия изоформ SIRT1 с транскрипционными факторами; зеленым цветом отмечены взаимодействия, которые подтверждаются на уровне экспрессии мРНК

Белок-белковые взаимодействия	CT[1]			Корреляции уровней мРНК	Коэффициент корреляции Спирмена (r)			
	SIRT1 V1	SIRT1 V2	SIRT1 V3		*SIRT1* общий	*SIRT1 V1*	*SIRT1 V2*	*SIRT1 V3*
PGC-1α	0.029	0.738	0.882	***PGC-1α***	0.437*	0.307	0.613*	0.717*
PPAR-α	0.642	0.963	0.953	***PPAR-α***	0.216	0.357*	0.630*	0.403
PPAR-γ	0.096	0.942	0.954	***PPAR-γ***	0.081	0.190	0.400	0.333

[1] CT – показатель вероятности белок-белкового взаимодействия, основанный на машинном обучении. Чем ближе CT к 1, тем выше вероятность взаимодействия; * $p < 0.05$

На уровне экспрессии мРНК были обнаружены положительные взаимосвязи ($p < 0.05$) между *PGC-1α* и *V2*; *PGC-1α* и *V3*, что согласуется с данными биоинформатического анализа. Однако белок-белковые взаимодействия между PPAR-α и цитоплазменными изоформами V2 и V3 подтверждаются на уровне экспрессии мРНК частично: *PPAR-α* положительно взаимосвязан с *V1* и *V2*, но не *V3*. Было предсказано взаимодействие между PPAR-γ и V2, V3, однако положительных корреляций между *PPAR-γ* и изоформами *SIRT1* в ходе исследования выявлено не было. Стоит отметить, что между *SIRT1* общим – канонической формой *SIRT1* – не было обнаружено положительной корреляции с мРНК *PPAR-α*. Из этого следует, что изучение только канонической формы *SIRT1* недостаточно (см. табл. 1).

Соотношение уровней мРНК изоформ *SIRT1* статистически не различалось между группами больных МС и здоровых доноров (рис. 1, а). Из всех изоформ только *V3* была повышена в печени больных МС (см. рис. 1, б). Несмотря на стабильную экспрессию референсного гена, у 16 доноров, вне зависимости от ИМТ, не было обнаружено экспрессии как *SIRT1* общего, так и его изоформ. У пяти доноров в обеих группах была обнаружена дискондартная экспрессия *SIRT1:* детектировался общий *SIRT1*, но не его изоформы.

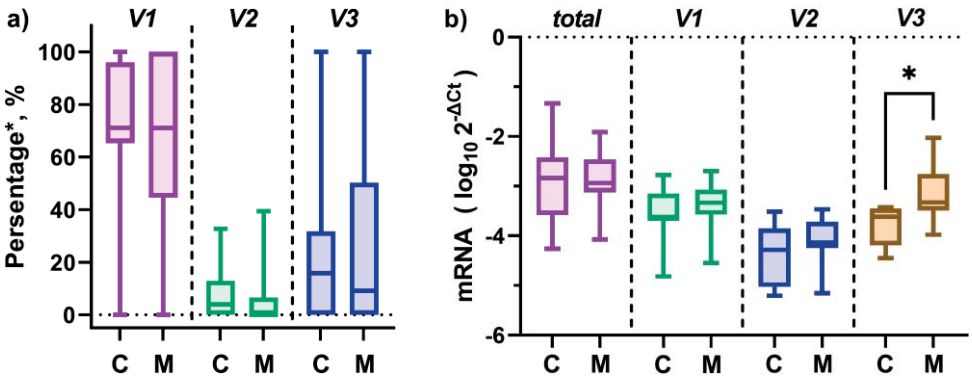


Рис. 1. Результаты анализа экспрессии изоформ в печени: а) соотношение изоформ, U-критерий Манна–Уитни, ($p < 0.05$); б) уровни экспрессии *SIRT1 общего, V1, V2, V3.* Т-критерий Уэлча, ($p < 0.05$); * $p < 0.05$

Выводы: Основываясь на результатах биоинформатического анализа и оценке уровня экспрессии мРНК изоформ *SIRT1* и перечисленных транскрипционных факторов, мы предполагаем, что изоформа V1 деацетилирует гистоны промоторов PPAR-α и PPAR-γ в ядре клетки, в то время как V2 и V3 воздействуют непосредственно на PGC-1α, PPAR-α в цитоплазме (рис. 2) [4]. Были предсказаны белок-белковые взаимодействия PPAR-γ и изоформ SIRT1, однако на уровне экспрессии мРНК положительных корреляций выявлено не было. Выявленные изменения могут быть опосредованы корепрессорным взаимодействием между SIRT1 и PPAR-γ: SIRT1 деацетилирует промотер PPAR-γ, дестабилизируя его, PPAR-γ подавляет экспрессию SIRT1 как транскрипционный фактор (см. рис. 2) [5].

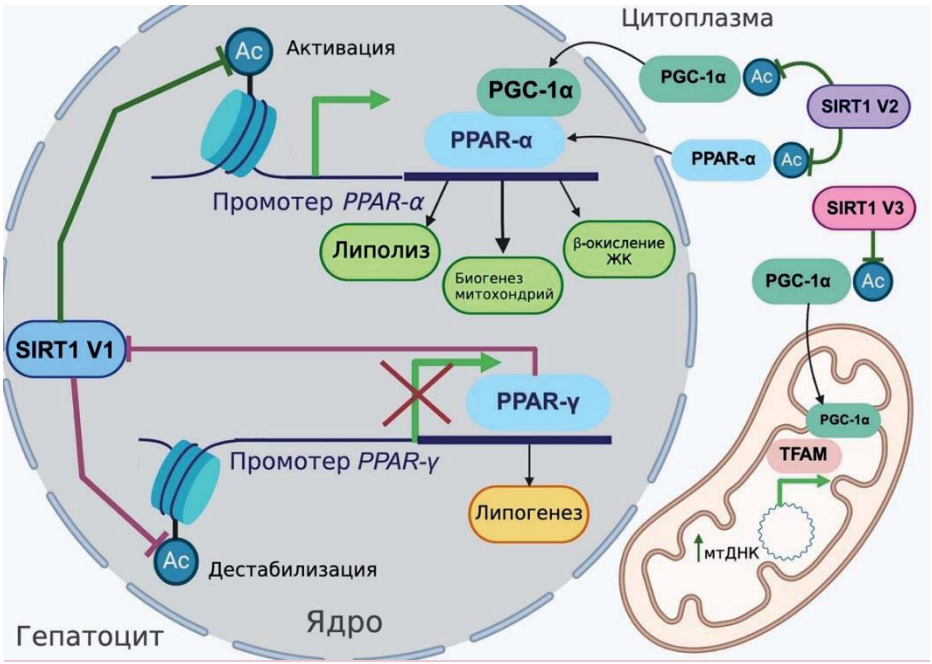


Рис. 2. Регуляция экспрессии транскрипционных факторов изоформами SIRT1 в печени

Соотношение уровней мРНК изоформ *SIRT1* в печени у контрольной группы и группы с МС не различалось, однако экспрессия *V3* была выше у больных МС. Таким образом, изоформы *SIRT1* могут различно модулировать биогенез митохондрий, липогенез и липолиз, что указывает на необходимость дальнейших исследований в этой области.

Финансирование: Исследование выполнено при финансовой поддержке Российского научного фонда (проект № 23-15-00061).

Splicing isoforms of SIRT1 in the liver of patients with metabolic syndrome – expectation/reality

Gorbacheva A.M.*, Voronova S.S., Bograya M.M., Vulf M.A., Litvinova L.S.
Immanuel Kant Baltic Federal University, Kaliningrad, Russia
** anmigorbacheva@gmail.com*

Key words: SIRT1, SIRT1 isoforms, metabolic syndrome

Motivation and Aim: Sirtuin 1 (SIRT1) is an NAD+/NADPH-dependent deacetylase involved in the modulation of lipogenesis and lipolysis as well as mitochondrial biogenesis by regulating the activity of the transcription factors PGC-1α, PPAR-α and PPAR-γ. SIRT1 has been extensively studied in context of metabolic syndrome (MetS), but its isoforms remain largely unexplored [1]. SIRT1 isoforms are generated by alternative splicing: isoform 1 (V1) has two nuclear localization signals, while the other two isoforms (V2 and V3) are localized in the cytoplasm. Therefore, *the aim* of current study was to perform a bioinformatic prediction of the protein-protein interactions (PPIs)

between the SIRT1 isoforms and the transcription factors PGC-1α, PPAR-α and PPAR-γ and to investigate the expression levels of the *SIRT1* isoforms and the listed transcription factors in the liver of patients with MetS.

Methods and Algorithms: PPIs were predicted using the PEPPI tool [2]. Amino acid sequences were used from UniProt: SIRT1 V1 (Q96EB6), SIRT1 V2 (E9PC49), SIRT1 V3 (B0QZ35), PGC-1α (Q9UBK2), PPAR-α (Q07869), PPAR-γ (P37231) [3].

Liver biopsies from 63 donors (mean age = 45 ± 8.48 years, 40 women and 23 men) were analyzed. The donors were divided into two groups: 1) control group (mean BMI = 19.8 ± 1.4 kg/m^2, 6 women, 8 men); 2) patients with MetS (mean BMI = 44.3 ± ± 12.3 kg/m^2, 34 women, 15 men), inclusion criteria: a history of at least two MetS components, fasting blood glucose >5.5 mmol/L and an atherogenic index >3. The expression level of the genes of interest was analyzed by quantitative RT-PCR using SYBR Green (*Evrogen*, Russia).

Results: All three SIRT1 isoforms were predicted to bind with the transcription factor PPAR-α. In addition, V2 and V3 bind to PGC-1α and PPAR-γ (Table 1).

At the level of mRNA expression, positive correlations ($p < 0.05$) were observed between *PGC-1α* and *V2*, *PGC-1α* and *V3*, which is consistent with the bioinformatic analysis data. However, PPIs between PPAR-α and cytoplasmic isoforms V2 and V3 are partially supported at the level of mRNA expression: *PPAR-α* is positively correlated with *V1* and *V2*, but not with *V3*. Interactions between PPAR-γ and V2, V3 were predicted, but no positive correlations were found between *PPAR-γ* and the *SIRT1* isoforms during the study. It is worth noting that no positive correlation was found between total *SIRT1* – the canonical form of *SIRT1* – and *PPAR-α*. This indicates that the study of only the canonical form of SIRT1 is insufficient (Table 1).

Table 1. Bioinformatic and experimental analysis of the interaction of SIRT1 isoforms with transcription factors; interactions confirmed at the mRNA expression level are highlighted in green

PPIs	CT[1]			Correlation of mRNA expression	Spearman correlation coefficient (r)			
	SIRT1 V1	SIRT1 V2	SIRT1 V3		*SIRT1* общий	*SIRT1 V1*	*SIRT1 V2*	*SIRT1 V3*
PGC-1α	0.029	0.738	0.882	***PGC-1α***	0.437*	0.307	0.613*	0.717*
PPAR-α	0.642	0.963	0.953	***PPAR-α***	0.216	0.357*	0.630*	0.403
PPAR-γ	0.096	0.942	0.954	***PPAR-γ***	0.081	0.190	0.400	0.333

[1] CT – a machine learning based indicator for the probability of PPI. The closer CT is to 1, the higher the probability of interaction. * $p < 0.05$

The ratio of mRNA levels of *SIRT1* isoforms did not differ statistically between the groups of patients with MetS and healthy donors (Fig. 1a). Of all isoforms, only *V3* was elevated in the livers of patients with MetS (Fig. 1b). Despite the stable expression of the reference gene, no expression of both total *SIRT1* and its isoforms was detected in 16 donors, regardless of BMI. Inconsistent expression of *SIRT1* was detected in 5 patients in both groups: total *SIRT1* was detected, but not its isoforms.

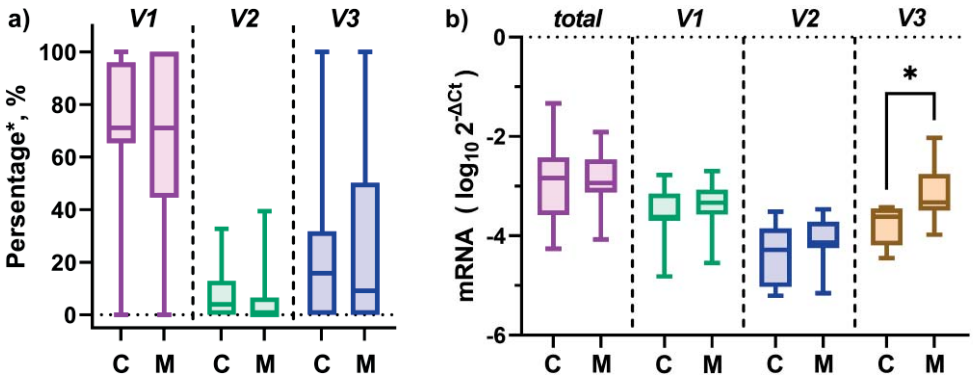


Fig. 1. Results of the analysis of the isoforms' expression in the liver: a) the ratio of isoforms, Mann–Whitney U test, ($p < 0.05$); b) the expression levels of total *SIRT1*, *V1*, *V2*, *V3*. Welch's t-test, ($p < 0.05$); * p-value < 0.05

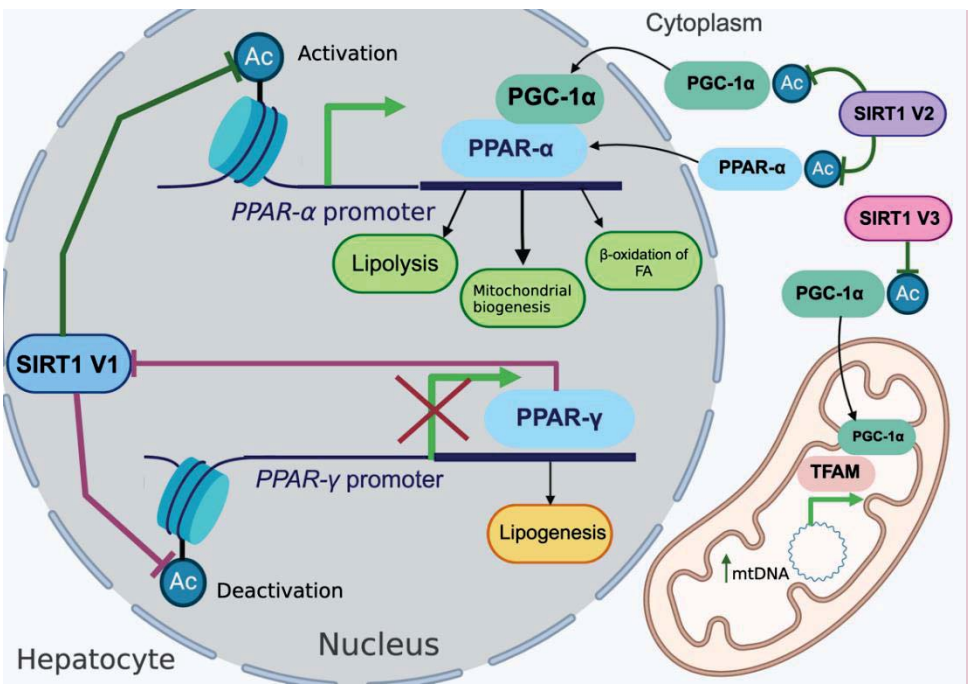


Fig. 2. Expression regulation of transcription factors by SIRT1 isoforms in the liver

Conclusion: Based on the results of bioinformatic analysis and evaluation of the mRNA expression levels of SIRT1 isoforms and the listed transcription factors, we suggest that isoform V1 deacetylates the histones of PPAR-α and PPAR-γ promoters in the cell nucleus, whereas V2 and V3 directly interact with PGC-1α and PPAR-α in the cytoplasm (Fig. 2) [4]. PPIs between PPAR-γ and SIRT1 isoforms were predicted, but no positive

correlations were observed at the level of mRNA expression. The observed changes could be mediated by a corepressor interaction between SIRT1 and PPAR-γ: SIRT1 deacetylates the PPAR-γ promoter and destabilizes it, while PPAR-γ represses SIRT1 expression as a transcription factor (Fig. 2) [5].

The ratio of mRNA levels of *SIRT1* isoforms in the liver did not differ between the control group and the group with MetS. However, the expression of V3 was higher in patients with MetS. Thus, SIRT1 isoforms may differentially modulate mitochondrial biogenesis, lipogenesis and lipolysis, indicating the need for further research in this area.

Funding: This research was funded by the Russian Science Foundation (Project No. 23-15-00061).

Список литературы/References

1. Zhang X., Ameer F.S., Azhar G., Wei J.Y. Alternative splicing increases sirtuin gene family diversity and modulates their subcellular localization and function. *Int J Mol Sci.* 2021;(2):473. doi 10.3390/ijms22020473
2. Bell E.W., Schwartz J.H., Freddolino P.L., Zhang Y. PEPPI: whole-proteome protein-protein interaction prediction through structure and sequence similarity, functional association, and machine learning. *J Mol Biol.* 2022;434(11):167530. doi 10.1016/j.jmb.2022.167530
3. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 2023;51(D1):D523-D553. doi 10.1093/nar/gkac1052
4. Islam H., Edgett B.A., Gurd B.J. Coordination of mitochondrial biogenesis by PGC-1α in human skeletal muscle: A re-evaluation. *Metabolism.* 2018;79:42-51. doi 10.1016/j.metabol.2017.11.001
5. Picard F., Kurtev M., Chung N. et al. Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR-γ. *Nature.* 2004;429(6993):771-776. doi 10.1038/nature02583

doi 10.18699/bgrs2024-9.1-06

Анализ молекулярных последовательностей гена CREBBP методом главных компонент (PCA-seq)

Гун Жуйхань [1]*, Ефимов В.[1, 2]

[1] *Новосибирский государственный университет, Новосибирск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

* *z.gun@g.nsu.ru*

Ключевые слова: crebbp; pca-seq

Ген CREBBP расположен в хромосоме 16p13.3, имеет 33 экзона и кодирует белок CBP из 2442 аминокислот [1]. Это коактиватор, который играет ключевую роль в эмбриональном развитии, контроле роста и стабильности внутренней среды, взаимодействуя с различными факторами транскрипции. Он может влиять на различные физиологические процессы и широко экспрессируется в тканях человека, включая семенники, костный мозг, яичники и эндометрий. Спектр мутаций CREBBP состоит из точечных вариантов (30–50 %) и делеций (~10 %) [2]. CBP экспрессируется повсеместно и, как было показано, обладает внутренней активностью гистон-ацетилтрансферазы (HAT), которая влияет на активность хроматина посредством модуляции нуклеосомных гистонов [3]. Кроме того, согласно исследованиям, точечные мутации в основном затрагивают домен HAT CBP, а потеря активности HAT достаточна, чтобы вызвать RSTS – наследственное заболевание, которое характеризуется комплексом врожденных аномалий. Хромосомные транслокации с участием этого гена также связаны с острым миелолейкозом.

Мы проанализировали взаимосвязь между тремя положениями каждого кодона гена и нуклеотидом (ACTG), и данные показали, что в целом C (цитозин) показал положительную корреляцию, а T (тимин) – отрицательную корреляцию, когда A(аденин) и G (гуанин) расположены в первой позиции кодона, корреляция мало чем отличается, но, когда они расположены во второй и третьей позициях соответственно, корреляция прямо противоположная.

Затем мы дополнительно проанализировали корреляцию между различными кодонами каждой аминокислоты, и данные показали, что корреляция глутаминовой кислоты, кодируемой CAG, выше, чем у других кодонов, кодирующих глутаминовую кислоту, поскольку число заболеваний, связанных с геном CREBBP, достигает 600. многие, включая HTT (болезнь Хантингтона). Основным патогенным кодоном HTT является cag. Аналогично, другие кодоны с более высокой корреляцией могут быть связаны с другими заболеваниями.

Поэтому мы выбрали ген CREBBP, используя новый статистический метод анализа PCA-Seq. Для достижения поставленной цели были определены следующие задачи: сначала с помощью метода главных компонент проанализировать нуклеотидные и аминокислотные последовательности гена CREBBP, а также изучить взаимосвязь между главными компонентами этих последовательностей и физическими и химическими свойствами аминокислот. Во-вторых, изучить связь между главными компонентами этих последовательностей и вторичной структурой белка.

PCA (Principal Component Analysis) – метод уменьшения размерности матриц данных. Суть этого метода заключается в построении ортогональной системы координатных осей (главных компонент), которая наиболее экономично описывает взаимное расположение объектов в евклидовом пространстве. Главные компоненты не коррелируют друг с другом и расположены в порядке убывания дисперсии. Изобретенный Карлом Пирсоном в 1901 году, он сейчас используется в различных областях для минимизации размеров данных при сохранении максимального количества информации. Для последовательностей PCA-Seq включает разрезание последовательности на фрагменты с использованием скользящего окна, расчет матрицы евклидовых расстояний для всех фрагментов и применение метода PCo для расчета главных компонент [4]. AAindex – это база данных числовых индексов, отражающая физические, химические и биологические свойства аминокислот и их пар. База данных AAindex широко используется в биоинформатических исследованиях и постоянно обновляется. Примеры включают предсказание субклеточной локализации белков, сайтов модификации и выравнивания последовательностей нескольких белков. База данных AAindex широко используется в исследованиях в области биоинформатики и постоянно обновляется [5].

Analysis of molecular sequences of the CREBBP gene by principal component analysis (PCA-seq)

Gong Ruihan[1]*, Efimov V.[1, 2]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

* *z.gun@g.nsu.ru*

Key words: crebbp; pca-seq

The CREBBP gene is located on chromosome 16p13.3, has 33 exons and encodes the CBP protein of 2442 amino acids [1]. It is a coactivator that plays a key role in embryonic development, growth control and stability of the internal environment by interacting with various transcription factors. It can influence various physiological processes and is widely expressed in human tissues, including testes, bone marrow, ovaries and endometrium. The spectrum of CREBBP mutations consists of point variants (30–50 %) and deletions (~10 %) [2]. CBP is expressed ubiquitously and has been shown to have intrinsic histone acetyltransferase (HAT) activity, which influences chromatin activity through modulation of nucleosomal histones [3]. Additionally, research suggests that point mutations primarily affect the HAT domain of CBP, and loss of HAT activity is sufficient to cause RSTS, an inherited disorder characterized by a complex of congenital abnormalities. Chromosomal translocations involving this gene are also associated with acute myeloid leukemia.

We analyzed the relationship between the three positions of each gene codon and the nucleotide (ACTG), and the data showed that overall, C (cytosine) showed a positive correlation and T (thymine) showed a negative correlation when A (adenine) and G (guanine) are located at the first codon position, the correlation is not much different, but when they are located at the second and third positions respectively, the correlation is exactly the opposite.

We then further analyzed the correlation between different codons of each amino acid, and the data showed that the correlation of CAG-encoded glutamic acid is higher than other glutamic acid-encoding codons as the number of diseases associated with the CREBBP gene reaches 600, including HTT. (Huntington's disease). The main pathogenic codon of HTT is cag. Likewise, other codons with higher correlations may be associated with other diseases.

Therefore, we selected the CREBBP gene using a new statistical analysis method, PCA-Seq. To achieve this goal, the following tasks were identified: first, using the principal component method, analyze the nucleotide and amino acid sequences of the CREBBP gene, and also study the relationship between the main components of these sequences and the physical and chemical properties of amino acids. Second, study the relationship between the main components of these sequences and the secondary structure of the protein.

PCA (Principal Component Analysis) is a method for reducing the dimension of data matrices. The essence of this method is to construct an orthogonal system of coordinate axes (principal components), which most economically describes the relative position of objects in Euclidean space. The principal components do not correlate with each other and are arranged in decreasing order of variance. Invented by Karl Pearson in 1901, it is now used in various fields to minimize the size of data while maintaining the maximum amount of information. For sequences, PCA-Seq involves cutting the sequence into fragments using a sliding window, calculating the Euclidean distance matrix for all fragments, and using the PCo method to calculate principal components [4]. AAindex is a database of numerical indexes reflecting the physical, chemical and biological properties of amino acids and their pairs. The AAindex database is widely used in bioinformatics research and is constantly updated. Examples include prediction of protein subcellular localization, modification sites, and sequence alignment of multiple proteins. The AAindex database is widely used in bioinformatics research and is constantly updated [5].

Список литературы/References

1. Pérez-Grijalba V. et al. New insights into genetic variant spectrum and genotype-phenotype correlations of Rubinstein-Taybi syndrome in 39 CREBBP-positive patients. *Mol Genet Genomic Med*. 2019;7(11):e972. doi 10.1002/mgg3.972
2. Rusconi D. et al. Characterization of 14 novel deletions underlying Rubinstein-Taybi syndrome: an update of the CREBBP deletion repertoire. *Hum Genet*. 2015;134(6):613-626. doi 10.1007/s00439-015-1542-9
3. Chan H.M., La Thangue N.B. p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J Cell Sci*. 2001;114(Pt. 13):2363-2373. doi 10.1242/jcs.114.13.2363
4. Efimov V.M., Efimov K.V., Kovaleva V.Y. Principal component analysis and its generalizations for any type of sequence (PCA-Seq). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2019;23(8):1032-1036. doi 10.18699/VJ19.584
5. Kawashima S., Pokarowski P., Pokarowska M., Kolinski A., Katayama T., Kanehisa M. AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res*. 2008;36(Database issue):D202-D205. doi 10.1093/nar/gkm998

doi 10.18699/bgrs2024-9.1-07

Представленность альтернативных изоформ РНК гена SLC4A11 в эндотелии роговицы здоровых людей и пациентов с эндотелиальной дистрофией роговицы Фукса

Ковалева П.А.[1]*, Котова Е.С.[1], Антонова О.П.[2], Малюгин Б.Е.[2], Ткаченко И.С.[2], Шарова Е.И.[1], Скородумова Л.О.[1]

[1] *ФНКЦ ФХМ имени ак. Ю.М. Лопухина ФМБА России, Москва, Россия*

[2] *НМИЦ «МНТК "Микрохирургия глаза" им. ак. С.Н. Федорова» МЗ РФ, Москва, Россия*

* *polina_a_kovaleva@mail.ru*

Ключевые слова: дистрофия Фукса; SLC4A11; роговица; 5’RACE

Мотивация и цель: Первичная дистрофия эндотелия роговицы Фукса (ДФ) является наследственным заболеванием глаза, характеризующимся снижением числа клеток эндотелия роговицы, выполняющих функцию поддержания водного баланса стромы роговицы. Известно, что ДФ является мультилокусным заболеванием. Наиболее распространенным каузальным вариантом (около 70 %) среди популяций европейского происхождения является экспансия тринуклеотидных повторов CTG18.1 в гене TCF4. По литературным данным часть случаев ДФ обусловлена вариантами в гене SLC4A11, белковый продукт которого является ионным транспортером и может вносить вклад в поддержание водного баланса стромы роговицы [1]. На данный момент большое количество литературных данных, связанных с функционированием белкового продукта SLC4A11 в эндотелии роговицы, было основано на данных о существовании трех основных РНК изоформ, отличающихся 5’-концевыми участками и кодируемых ими белковых продуктов. В большинстве статей на основе различных данных и предположений в качестве основной изоформы РНК называлась изоформа 2, а в качестве основной изоформы белка – ее продукт. Также в некоторых работах исследования проводились на изоформах РНК 3 и 1 и их белковых продуктах [2, 3]. Кроме того были получены свидетельства, что основной формой белка SLC4A11 в эндотелии роговицы является укороченный с N-конца полипептид (далее-v2short), который считают результатом старта трансляции на изоформе РНК 2 не с первого, а с третьего старт-кодона, находящегося в той же рамке считывания, что и первый [4]. Таким образом, целью нашей работы стал анализ представленности альтернативных изоформ РНК гена SLC4A11 в эндотелии роговицы здоровых людей и пациентов с эндотелиальной дистрофией роговицы Фукса.

Методы и алгоритмы: Аннотационный gtf-файл с полным набором транскриптов SLC4A11 формировался на основе представленных транскриптов в базе данных Refseq и проекта Gencode с дополнением транскриптов, полученных de novo StringTie (2.2.1). Также для транскриптов гена SLC4A11, полученных StringTie, проводился сравнительный анализ аннотации пакетом GffCompare (0.12.6) и для предсказания открытых рамок считывания (ORF) и нахождения потенциальных кодирующих последовательностей в транскриптомных данных использовали пакет TransDecoder (5.5.0). Полученный аннотационный gtf-файл использовали

для оценки уровня экспрессии генов и изоформ с помощью пакета Salmon (1.3.0) для двух сетов транскриптомных данных, представленных образцами эндотелия роговицы доноров и пациентов с ДФ [5, 6]. После проводился анализ дифференциальной экспрессии гена SLC4A11 с помощью пакета EdgeR (2.0.7). Для подтверждения данных об уровне экспрессии изоформ, полученных биоинформатически, было проведено выявление 5’-концевых участков транскриптов, образующихся в эндотелии роговицы человека методом быстрой амплификации 5’-концевых фрагментов 5’RACE (Rapid Amplification of cDNA Ends). При этом из образцов эндотелия роговицы 5 здоровых доноров и 5 пациентов с ДФ была выделена РНК, которая далее использовалась для получения наборов ампликонов, соответствующих 5’-концевым участкам транскриптов SLC4A11, содержавшихся в этих образцах. Обратная транскрипция с переключением матрицы проводилась набором SMARTer RACE 5'/3' Kit (Takara). На матрице полученных образцов кДНК проводились 2 раунда нестед ПЦР с ПЦР-супрессией. После этого проводилось секвенирование полученных ампликонов по Сэнгеру и на Illumina MiSeq для определения и аннотации транскриптов, которые экспрессируются в эндотелии роговицы.

Результаты: Мы провели предварительный анализ покрытия прочтениями локуса гена SLC4A11, используя опубликованные транскриптомы (тотальные и поли-А) эндотелия здоровых доноров [5, 7] и наибольшая плотность прочтений соответствовала 5’-области РНК-изоформы, представленной в базе данных RefSeq, как NM_001400277, или изоформа 6, которая не соответствовала ни одной из трех основных изоформ SLC4A11, фигурирующих в литературных данных. Также высокая плотность прочтений приходилась на 5’-область третьей изоформы. При этом покрытие прочтениями 5’-области, соответствующей 2 и 1 изоформам, было довольно низким.

Затем была проведена оценка уровня экспрессии изоформ SLC4A11 для двух сетов транскриптомных данных, представленных образцами эндотелия роговицы доноров и пациентов с ДФ [5, 6].

De novo сборка транскриптомов гена SLC4A11 выявила две дополнительные изоформы MSTRG.58852.10 и MSTRG.58852.21. Пакетом TransDecoder для этих MSTRG.58852.10 и MSTRG.58852.21 были выявлены белок-кодирующие последовательности. Однако их кодирующая роль оценивается нами как мало значимая, так как MSTRG.58852.21 кодирует довольно короткий полипептид, а у MSTRG.58852.10 после стоп-кодона идут экзон-экзонные сочленения, это указывает на то, что для данной изоформы в клетке скорее всего реализуется механизм nonsense-mediated decay.

При оценке представленности изоформ в двух сетах данных эндотелия роговицы наиболее представленной изоформой оказалась изоформа 3, второй по представленности – изоформа 6 (NM_001400277), которая не соответствовала ни одной из трех основных изоформ SLC4A11. Третьей по представленности оказалась некодирующая изоформа NR_174471. Высокая представленность изоформы 6 интересна тем, что она кодирует v2short, но при этом его старт трансляции является первым ATG-кодоном данной мРНК. Представленность изоформ 1 и 2 была оценена как довольно низкая.

Секвенирование по Сэнгеру продуктов, полученных методом быстрой амплификации 5’-концевых фрагментов целевых транскриптов гена SLC4A11, показало, что образцы соответствовали 5’-концевому участку изоформы 6 и

некодирующей изоформы NR_174471. Предварительный анализ NGS-секвенирования одного из образцов, показал, что в нем наиболее представленной является 5'-конец изоформы 6, гораздо менее представлен конец изоформы 3, практически нет прочтений, соответствующих изоформам 1 и 2 (1 прочтение) и есть небольшое количество прочтений, соответствующих сплайсированным транскриптам, начинающимся внутри второго интрона 1-3 и 6 изоформ, о промоторной активности этого участка генома пока не обнаружено никаких данных.
Выводы: Мы получили свидетельства преимущественной экспрессии РНК изоформ SLC4A11 3 и 6 в эндотелии роговицы и довольно низкой представленности в нем изоформы 2. Эти данные не только помогут уточнить информацию о преимущественно образующихся в эндотелии роговицы белковых изоформах SLC4A11, но также охарактеризовать функционирование промоторного региона SLC4A11в эндотелии роговицы.
Финансирование: Исследование поддержано грантом Российского научного фонда (№ 22-75-10157).

Study of the SLC4A11 alternative RNA isoforms expression in the corneal endothelium of healthy individuals and patients with Fuchs corneal endothelial dystrophy

Kovaleva P.A.[1]*, Kotova E.S.[1], Antonova O.P.[2], Malyugin B.E.[2], Tkachenko I.S.[2], Sharova E.I.[1], Skorodumova L.O.[1]

[1] *Lopukhin FRCC PCM, Moscow, Russia*

[2] *The S. Fyodorov Eye Microsurgery Federal State Institution, Moscow, Russia*

* *polina_a_kovaleva@mail.ru*

Key words: FECD; SLC4A11; cornea; 5'RACE

Motivation and Aim: Primary Fuchs' endothelial corneal dystrophy (FECD) is an inherited eye disease characterized by a reduction in the number of corneal endothelial cells, which are responsible for maintaining the water balance in the corneal stroma. FECD is known to be a multilocus disease. The most common causative variant (approximately 70 %) in populations of European descent is the expansion of the trinucleotide repeat CTG18.1 in the TCF4 gene. According to the literature, a proportion of FECD cases are caused by variants in the SLC4A11 gene, the protein product of which is an ion transporter and may contribute to maintaining water balance in the corneal stroma [1]. To date, much of the literature on the function of the SLC4A11 protein product in the corneal endothelium has been based on the existence of three major RNA isoforms that differ in their 5'-ends and the protein products encoded by them. Most papers, based on different data and assumptions, identified isoform 2 as the major RNA isoform and its product as the major protein isoform. Some papers have also looked at RNA isoforms 3 and 1 and their protein products [2, 3]. In addition, it has been shown that the major form of SLC4A11 protein in the corneal endothelium is an N-short polypeptide (hereafter referred to as v2short), which is considered to be the result of translation start of RNA isoform 2 not from the first but from the third start codon, which is in the same reading frame as the first [4]. Therefore, the aim of our work was to analyze

the expression of alternative RNA isoforms of the SLC4A11 gene in the corneal endothelium of healthy subjects and patients with FECD.

Methods and Algorithms: An annotation gtf-file with the complete set of SLC4A11 transcripts was generated based on the submitted transcripts in the Refseq database and the Gencode project with the addition of transcripts obtained de novo by StringTie (2.2.1). Similarly, for the SLC4A11 gene transcripts obtained by StringTie, the annotation was compared using the GffCompare package (0.12.6) and the TransDecoder package (5.5.0) was used to predict open reading frames (ORFs) and find potential coding sequences in the transcriptome data. The resulting annotated gtf-file were used to assess gene expression levels and isoforms using the Salmon package (1.3.0) for two sets of transcriptomic data represented by corneal endothelial samples from donors and patients with FECD [5, 6]. Differential expression of the SLC4A11 gene was analyzed using the EdgeR package (2.0.7). To confirm the bioinformatic data on isoform expression levels, the 5'-end regions of transcripts produced in human corneal endothelium were identified by rapid amplification of 5'-end fragments using 5'RACE (Rapid Amplification of cDNA Ends). RNA was isolated from corneal endothelial samples of 5 healthy donors and 5 patients with FECD and used to obtain sets of amplicons corresponding to 5'-end fragments of SLC4A11 transcripts present in these samples. Reverse transcription with matrix switching was performed using the SMARTer RACE 5′/3′ Kit (Takara). 2 rounds of nested PCR with PCR suppression were performed on the matrix of the obtained cDNA samples. The resulting amplicons were then sequenced by Sanger approach and on Illumina MiSeq to identify and annotate the transcripts expressed in the corneal endothelium.

Results: We performed a preliminary analysis of read coverage of the SLC4A11 gene locus using published transcriptomes (total and poly-A) of healthy donor endothelium [5, 7] and the highest density of reads corresponded to the 5' region of the RNA isoform represented in the RefSeq database as NM_001400277 or isoform 6, which does not correspond to any of the three major SLC4A11 isoforms in the literature. There was also a high density of reads in the 5' region of the third isoform. However, the coverage of reads in the 5' region corresponding to isoforms 2 and 1 was rather low.

We then evaluated the expression level of SLC4A11 isoforms for two sets of transcriptomic data represented by corneal endothelial samples from donors and patients with FECD [5, 6].

The expression level of SLC4A11 isoforms was then evaluated for two sets of transcriptomic data represented by corneal endothelial samples from donors and FECD patients [5, 6].

De novo assembly of the SLC4A11 gene transcriptome revealed two additional isoforms, MSTRG.58852.10 and MSTRG.58852.21. The protein coding sequences for these MSTRG.58852.10 and MSTRG.58852.21 were identified using the TransDecoder package. However, we consider their coding role to be of little importance since MSTRG.58852.21 encodes a rather short polypeptide and MSTRG.58852.10 has exon-exon junctions after the stop codon, indicating that the nonsense-mediated decay mechanism is most likely realized in the cell for this isoform.

When evaluating the representation of isoforms in two sets of corneal endothelial data, isoform 3 was the most represented isoform, the second most represented isoform was isoform 6 (NM_001400277), which did not correspond to any of the three major isoforms of SLC4A11. The third most abundant isoform was the non-coding isoform NR_174471. The high representation of isoform 6 is interesting because it encodes

v2short, but its translation start is the first ATG codon of this mRNA. The representation of isoforms 1 and 2 was estimated to be rather low.
Sanger sequencing of the products obtained by rapid amplification of 5'-end fragments of the target transcripts of the SLC4A11 gene showed that the samples corresponded to the 5'-end of isoform 6 and the non-coding isoform NR_174471. Preliminary NGS sequencing analysis of one of the samples showed that the 5' end of isoform 6 is most abundant, the end of isoform 3 is much less abundant, there are virtually no reads corresponding to isoforms 1 and 2 (1 read), and there are a small number of reads corresponding to spliced transcripts starting within the second intron of isoforms 1-3 and 6. No data on the promoter activity of this genomic region have been found.
Conclusion: We obtained evidence for the predominant RNA expression of SLC4A11 isoforms 3 and 6 in the corneal endothelium and a rather low representation of isoform 2. These data will not only help to clarify the information about the protein isoforms of SLC4A11 predominantly produced in the corneal endothelium, but also to characterize the function of the promoter region of SLC4A11 in the corneal endothelium.
Funding: The study is supported by Russian Science Foundation, grant No. 22-75-10157.

Список литературы/References

1. Tsedilina T.R., Sharova E., Iakovets V., Skorodumova L.O. Systematic review of SLC4A11, ZEB1, LOXHD1, and AGBL1 variants in the development of Fuchs' endothelial corneal dystrophy. *Front Med*. 2023;10:1153122. doi 10.3389/fmed.2023.1153122
2. Hara S., Tsujikawa M., Kawasaki S., Nishida K. Homeostasis of SLC4A11 protein is mediated by endoplasmic reticulum-associated degradation. *Experimental Eye Res*. 2019;188:107782. doi 10.1016/j.exer.2019.107782
3. Kao L., Azimov R., Shao X.M. et al. Multifunctional ion transport properties of human SLC4A11: comparison of the SLC4A11-B and SLC4A11-C variants. *Am J Physiol Cell Physiol*. 2016;311(5):C820-C830. doi 10.1152/ajpcell.00233.2016
4. Malhotra D., Loganathan S.K., Chiu A.M., Lukowski C.M., Casey J.R. Human corneal expression of SLC4A11, a gene mutated in endothelial corneal dystrophies. *Sci Rep*. 2019;9(1):9681. doi 10.1038/s41598-019-46094-y
5. Nikitina A.S., Belodedova A.V., Malyugin B.E. et al. Dataset on transcriptome profiling of corneal endothelium from patients with Fuchs endothelial corneal dystrophy. *Data Brief*. 2019;25:104047. doi 10.1016/j.dib.2019.104047
6. Chu Y., Hu J., Liang H. et al. Analyzing pre-symptomatic tissue to gain insights into the molecular and mechanistic origins of late-onset degenerative trinucleotide repeat disease. *Nucleic Acids Res*. 2020;48(12):6740-6758. doi 10.1093/nar/gkaa422
7. Chen Y., Huang K., Nakatsu M.N., Xue Z., Deng S.X., Fan G. Identification of novel molecular markers through transcriptomic analysis in human fetal and adult corneal endothelial cells. *Hum Mol Genet*. 2013;22(7):1271-1279. doi 10.1093/hmg/dds527

doi 10.18699/bgrs2024-9.1-08

Нанопоровое секвенирование конкатемерных библиотек вариабельных областей TCR из опухолевой ткани колоректального рака

Котова Е.С.[1]*, Петрова Т.В.[1], Бабенко В.В.[1], Болдырева Д.И.[1], Гарманова Т.Н.[2], Лазарев В.Н.[1], Шарова Е.И.[1]

[1] *ФНКЦ ФХМ имени академика Ю.М. Лопухина ФМБА России, Москва, Россия*

[2] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

* *potamanthus@gmail.com*

Ключевые слова: нанопоровое секвенирование; конкатемеры; RCA; TCR; колоректальный рак

Мотивация и цель: Секвенирование нового поколения (NGS-Next Generation Sequencing) позволило провести множество важных исследований в области молекулярной онкологии. Важно отметить, что методы NGS все еще находятся в стадии разработки и имеют свои ограничения. Секвенаторы с высоким качеством прочтений не могут производить прочтения достаточной длины, а секвенаторы с длинными прочтениями не могут обеспечить прочтения высокой точности с приемлемым выходом. Нанопоровое секвенирование конкатемерных библиотек с получением консенсуса мономерных последовательностей позволяет получать достаточно длинные считывания с более высокой точностью [1, 2]. Мы предлагаем новый метод получения конкатемерных библиотек ампликонов без использования праймеров на этапе получения конкатемеров путем амплификации по типу катящегося кольца (RCA-Rolling Circle Amplification), для уменьшения доли химерных молекул. Мы провели предварительное тестирование разработанного протокола.

Методы и алгоритмы: С помощью набора Native Barcoding Kit 24 V14 были получены шесть баркодированных библиотек, соответствующих вариабельным областям альфа-цепи Т-клеточного рецептора (TCRA – T-cell receptor alpha chain) и бэта-цепи Т-клеточного рецептора (TCRB – T-cell receptor alpha chain) из трех образцов РНК опухолей колоректального рака. Каждый мономер в конкатемерной молекуле нес уникальный молекулярный индекс (УМИ) длиной 17 п.о. для обнаружения и фильтрации химерных прочтений. Секвенирование проводили на проточной ячейке R10.4. Полученные данные анализировали с помощью пакетов Bowtie2, MAFFT, MiXCR и собственных скриптов, написанных на bash, питоне 3 и R.

Результаты: Для библиотек вариабельных участков TCRA было получено 3954-6668 прочтений, TCRB – 4013-6427. Медианное количество мономеров для всех библиотек составило 7; 25 % прочтений содержали более 13–14 мономерных единиц, а 10 % – более 22–24. Консенсусные последовательности были получены для каждого конкатемера и проанализированы с помощью программы MiXCR. Мы идентифицировали 33–69 и 29–176 клонов TCRA и TCRB соответственно. Клонов, которым соответствовали, как минимум, 3 прочтения, было 3-21 (TCRA) и 4-14 (TCRB). Количество считываний для наиболее представленных клонов составило 31 (TCRA) и 22 (TCRB).

Выводы: Применение метода RCA без использования праймеров с последующим нанопоровым секвенированием позволило успешно реконструировать вариабельные области TCRA и TCRB полной длины на основе нуклеотидных последовательностей мономеров, несущих УМИ. Первичные данные свидетельствуют о том, что это перспективный подход к полноразмерному секвенированию и высокопроизводительному профилированию гетерогенных РНК с высоким уровнем сходства, к примеру, РНК TCR, BCR и ретроэлементов в опухолевых образцах.
Финансирование: Данная публикация выполнена в рамках государственного задания «Персональный онкопрепарат», номер государственного учета НИОКТР 124031200002-3.

Nanopore concatemeric sequencing of TCR variable regions from colorectal tumors

Kotova E.S.[1]*, Petrova T.V.[1], Babenko V.V.[1], Boldyreva D.I.[1], Garmanova T.N.[2], Lazarev V.N.[1], Sharova E.I.[1]

[1] *Lopukhin FRCC PCM, Moscow, Russia*

[2] *Moscow State University, Moscow, Russia, Moscow, Russia*

* *potamanthus@gmail.com*

Key words: nanopore sequencing; concatemers; RCA; TCR; colorectal cancer

Motivation and Aim: Next-generation sequencing has enabled many important studies in molecular oncology. It is important to note that this technique is still in development and has limitations. Sequencers with high read quality cannot provide reads that are long enough, and long-read sequencers cannot provide reads that are accurate enough with acceptable yields. Nanopore sequencing of concatemeric libraries, with the derivation of monomeric sequence consensus, allows for the obtainment of sufficiently long reads with higher accuracy [1, 2]. We propose a new method for obtaining concatemeric libraries of amplicons without primers during the concatemer production stage by using rolling circle amplification (RCA) to decrease the portion of chimeric molecules. We conducted a preliminary test of the developed protocol.
Methods and Algorithms: Six barcoded libraries were created using the Native Barcoding Kit 24 V14, corresponding to the variable regions of TCRA (T-cell receptor alpha chain) and TCRB (T-cell receptor beta chain) from three RNA samples of colorectal cancer tumors. Each monomer in the concatemeric molecule was labeled with a 17 bp unique molecular identifier (UMI) to detect and filter out chimeric reads. Sequencing was performed using the R10.4 flow cell. We analyzed obtained data using packages Bowtie2, MAFFT, MiXCR and our own bash-, python3- and R-scripts.
Results: The TCRA libraries yielded 3954-6668 reads, while the TCRB libraries yielded 4013-6427 reads. The median number of monomeric units for all libraries was 7; 25 % of the reads contained more than 13-14 monomeric units, and 10 % contained more than 22-24. Consensus sequences were obtained for each concatemer and analyzed using the Mixcr software. We identified 33-69 and 29-176 TCRA and TCRB clones respectively. The clones that matched at least 3 reads were 3-21 (TCRA) and 4-14 (TCRB). The number of reads for the most represented clones was 31 (TCRA) and 22 (TCRB).

Conclusion: The entire length of the TCRA and TCRB variable regions were successfully reconstructed using UMI-labeled monomers through a primerless RCA approach and nanopore sequencing. The initial data suggests that this approach is very promising for whole length sequencing and high throughput profiling of heterogeneous RNAs with high similarity, such as TCR, BCR, and retroelements in tumor samples.
Funding: This publication was carried out with state funding for the «Personalnyi onkopreparat» project, state registration number of R&D 124031200002-3.

Список литературы/References

1 Li C. et al. INC-Seq: accurate single molecule reads using nanopore sequencing. *Gigascience*. 2016;5(1):34
2 Volden R. et al. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc Natl Acad Sci USA*. 2018;115(39):9726-9731

doi 10.18699/bgrs2024-9.1-09

Современная оценка нкРНК как биомаркеров прогноза при светлоклеточном почечно-клеточном раке

Логинов В.И.[1, 2]*, Бурденный А.М.[1, 3], Пронина И.В.[1], Лукина С.С.[1], Брага Э.А.[1, 2]

[1] *Научно-исследовательский институт общей патологии и патофизиологии, Москва, Россия*

[2] *Медико-генетический научный центр им. академика Н.П. Бочкова, Москва, Россия*

[3] *Институт биохимической физики им. Н.М. Эмануэля РАН, Москва, Россия*

* *loginov7w@gmail.com*

Ключевые слова: мРНК; миРНК; днРНК; светлоклеточный почечноклеточный рак; скПКР

Мотивация и цель: Ключевую роль в развитии и прогрессии опухолей играют эпигенетические изменения. В частности, результаты многочисленных исследований однозначно подтверждают вовлеченность в онкогенез аномального метилирования ДНК. Региональное гиперметилирование промоторных CpG-островков ряда генов (белок-кодирующих, микроРНК, длинных некодирующих РНК и др.) – еще одно распространенное явление при опухолевом процессе. Результатом таких эпигенетических изменений может стать ингибирование транскрипции генов-супрессоров, обеспечивающих в нормальных клетках защиту от злокачественной трансформации [1]. Катализаторами добавления метильных групп служат ферменты из семейства ДНК-метилтрансфераз (DNMTs), в то время как за удаление модификации главным образом ответственны диоксигеназы TET [2]. Нарушения в работе этих ферментов, обусловленные мутациями, эпигенетическими событиями или иными причинами, могут быть ассоциированы с развитием злокачественных новообразований.

Целью настоящей работы был поиск возможных осей днРНК/миРНК/мРНК, участвующих в регуляции некоторых генов, ответственных за механизмы метилирования и деметилирования ДНК, при скПКР.

Методы и алгоритмы: Биоинформатический анализ проводился с использованием открытых баз данных входящих в базу NCBI (https://www.ncbi.nlm.nih.gov/): PubMed, Gene и др. В работе также были использованы Lnc2Cancer 2.0 (www.bio-bigdata.net/lnc2cancer), GEPIA2.0 (http://gepia2.cancer-pku.cn/#index), miRCode (http://www.mircode.org/), miRWalk 2.0 (http://mirwalk.umm.uni-heidelberg.de/), TargetScan 8.0. (https://www.targetscan.org/vert_80/) и The Genotype-Tissue Expression (GTEx, https://www.gtexportal.org/home/).

Результаты: Отбор днРНК, участвующих в прогрессии эпителиальных форм рака, в том числе скПКР, проводился на основании анализа данных литературы PubMed с использованием ключевых слов по формуле: “human” AND ((“long non-coding RNA”) OR (“lncRNA”) OR (“lincRNA”) OR (“long intergenic non-protein coding RNA”) OR (“long non-protein-coding RNA”)) AND “cancer”. Критерием отбора являлось то, что исследования должны были быть оригинальными и в них оценивалась экспрессия днРНК в опухолевой ткани эпителиальных опухолей, в том числе и при скПКР. Критерии исключения были следующими: публикации написаны на языке, отличном от английского; статьи относятся к категориям: отчеты о случаях заболевания, комментарии, письма, обзоры или метаанализы;

другие заболевания; исследования не оценивали роль днРНК в диагностике, прогнозе и ответе на лечение. Связь днРНК с онкозаболеваниями также была оценена с использованием баз данных LncRNADisease 2.0 (www.cuilab.cn/lncrnadisease) и lncATLAS (https://lncatlas.crg.eu/). Транскриптомный ландшафт днРНК был идентифицирован с помощью Lnc2Cancer 2.0. По результатам анализа было отобрано около 400 днРНК, изменяющих экспрессию в эпителиальных опухолях разных локализаций.

С привлечением базы данных GEPIA2.0 [3] был проведен интерактивный анализ профиля экспрессии отобранных генов при скПКР. Были отобраны гены с дифференциальной экспрессией в опухолевой ткани. В результате было отобрано около 150 днРНК, изменяющих экспрессию в опухолевой ткани более чем в 2 раза. Следующим этапом анализа было выявление наличия CpG островков в области не далее 2000 п.н. от 5' конца гена или старта транскрипции по базе Epigenomics, входящей в пакет NCBI. В результате поиска было отобрано 50 днРНК. Характеристики представленности днРНК и мРНК в нормальной ткани получены с помощью ресурса GTEx.

Одним из основных этапов биоинформатического анализа был поиск миРНК, взаимодействующих как с отобранными днРНК, так и белок-кодирующими генами (*TET1, TET2, TET3, TDG, DNMT3A, DNMT3B* и *DNMT1*). Скрининг был основан на анализе данных полногемного секвенирования NCBI GEO и определении процента комплементарных пар с помощью локального выравнивания с использованием алгоритма Смита-Ватермана. Затем проверяли отобранные миРНК по известным базам данных (miRCode, miRWalk 2.0 и TargetScan 8.0). Отбирали миРНК минимум с двумя сайтами взаимодействий. Минимальная свободная энергия (ΔG) <-15 ккал/моль была установлена в качестве порога взаимодействия. Рассматривались дуплексы миРНК/мРНК и миРНК/днРНК со спариванием оснований по меньшей мере из 7-8 нуклеотидов в затравочной последовательности миРНК. В результате, для дальнейшего использования были отобраны miR-124-3p, miR-129-5p и miR-125b-5p, негативно коррелирующие как с экспрессией мРНК белок-кодирующих генов-мишеней, так и днРНК.

В дальнейшем потребуется проверка полученных данных на репрезентативной выборке клинических образцов с последующей валидацией на культурах клеток.

Выводы: В результате комплексного анализа отобрано 17днРНК (*OIP5-AS1, FAM222A-AS1, GAS5, HAND2-AS1, LINC00339, LINC00886, MALAT1, MEG3, NCK1-DT, NEAT1, SNHG1, SNHG12, TINCR, TP73-AS1, XIST, SNHG6, ADAMTS9-AS2*), 3 миРНК (miR-124-3p, miR-129-5p, miR-125b-5p) и 7 белок-кодирующих генов (*TET1, TET2, TET3, TDG, DNMT3A, DNMT3B* и *DNMT1*). Проведена оценка вероятности их взаимодействия. Получены теоретические данные о комбинации триплетов (осей), возможно участвующих в развитии скПКР.

Финансирование: Исследование поддержано государственным заданием FGFU-2022-0007.

Modern assessment of ncRNAs as biomarkers of prognosis in clear cell renal cell carcinoma

Loginov V.I.[1, 2]*, Burdennyy A.M.[1, 3], Pronina I.V.[1], Lukina S.S.[1], Braga E.A.[1, 2]

[1] *Institute of General Pathology and Pathophysiology, Moscow, Russia*

[2] *Bochkov Research Centre for Medical Genetics, Moscow, Russia*

[3] *Emanuel Institute for Biochemical Physics, Moscow, Russia*

* *loginov7w@gmail.com*

Key words: mRNA; miRNA; lncRNA; clear cell renal cell carcinoma; ccRCC

Motivation and Aim: Epigenetic changes play a key role in the development and progression of tumors. In particular, the results of numerous studies unequivocally confirm the involvement of abnormal DNA methylation in oncogenesis. Regional CpG islands promoter hypermethylation of a number of genes (including genes coding proteins, microRNAs, long non-coding RNAs, etc.) is another common phenomenon in the tumor process. The result of these epigenetic changes may be the inhibition of the transcription of suppressor genes that protect normal cells from malignant transformation [1]. Enzymes from the DNA methyltransferase (DNMT) family catalyze the transfer of methyl groups, while TET dioxygenases remove this modification [2]. Disorders in the functioning of these enzymes induced by mutations, epigenetic events or other causes may be associated with the development of malignant neoplasms.

The aim of this work was to search for possible lncRNA/miRNA/mRNA axes involved in the regulation of some genes responsible for the mechanisms of DNA methylation and demethylation in ccRCC.

Methods and Algorithms: Bioinformatic analysis was carried out using open-access databases included in the NCBI database (https://www.ncbi.nlm.nih.gov/): PubMed, Gene, etc. Lnc2Cancer 2.0 (www.bio-bigdata.net/lnc2cancer), GEPIA2.0 (http://gepia2.cancer-pku.cn/#index), miRCode (http://www.mircode.org/), miRWalk 2.0 (http://mirwalk.umm.uni-heidelberg.de/), TargetScan 8.0. (https://www.targetscan.org/vert_80/) and The Genotype-Tissue Expression (GTEx, https://www.gtexportal.org/home/) was also used in the work.

Results: The selection of lncRNAs involved in the progression of epithelial forms of cancer, including ccRCC, was carried out based on the analysis of PubMed publication data using keywords according to the formula: “human” AND ((“long non-coding RNA”) OR (“lncRNA”) OR (“lincRNA”) OR (“long intergenic non-protein coding RNA”) OR (“long non-protein-coding RNA”)) AND “cancer”. The selection criteria were as follows: the studies had to be original and they evaluated the expression of lncRNA in the cancer tissue of epithelial tumors, including in ccRCC. The exclusion criteria were as follows: publications were written in a language other than English; articles belong to the categories: case reports, comments, letters, reviews or meta-analyses; other diseases; studies did not evaluate the role of lncRNA in diagnosis, prognosis and response to treatment. The association of lncRNA with cancer was also evaluated using LncRNADisease 2.0 databases (www.cuilab.cn/lncrnadisease) and lncATLAS (https://lncatlas.crg.eu/). The transcriptomic landscape of lncRNA was identified using Lnc2Cancer 2.0. According to the results of the analysis, about

400 lncRNAs were selected that change expression in epithelial tumors of different localizations.

Using the GEPIA2.0 database [3] an interactive analysis of the expression profile of the selected genes was performed in ccRCC. Genes with differential expression in tumor tissue were selected. As a result, about 150 lncRNAs were selected that change expression in tumor tissue by more than 2 times.

The next stage of the analysis was to identify the presence of CpG islands in the region no further than 2000 bp from the 5' end of the gene or the start of transcription according to the Epigenomics database included in the NCBI package. As a result of the search, 50 lncRNAs were selected. The characteristics of the representation of lncRNA and mRNA in normal tissue were obtained using the GTEx resource.

One of the main stages of bioinformatic analysis was the search for miRNAs interacting with both selected lncRNAs and protein-coding genes (*TET1, TET2, TET3, TDG, DNMT3A, DNMT3B* and *DNMT1*). The screening was based on the analysis of NCBI GEO full-system sequencing data and the determination of the percentage of complementary pairs using local alignment using the Smith-Waterman algorithm. Then the selected miRNAs were checked against well-known databases (miRCode, miRWalk 2.0 and TargetScan 8.0). miRNAs with at least two interaction sites were selected. The minimum free energy (ΔG) <-15 kcal/mol was set as the interaction threshold. Duplexes of miRNA/mRNA and miRNA/lncRNA with base pairing of at least 7-8 nucleotides in the miRNA seed sequence were considered. As a result, miR-124-3p, miR-129-5p and miR-125b-5p were selected for further use, negatively correlating with both mRNA expression of protein-coding target genes and lncRNA.

In the future, it will be necessary to verify the data obtained on a representative sample of clinical samples with subsequent validation on cell cultures.

Conclusion: As a result of the complex analysis, 16 lncRNAs (OIP5-AS1, FAM222A-AS1, GAS5, HAND2-AS1, LINC00339, LINC00886, MALAT1, MEG3, NCK1-DT, NEAT1, SNHG1, SNHG12, TINCR, TP73-AS1, XIST, SNHG6, ADAMTS9-AS2), 3 miRNAs (miR-124-3p, miR-129-5p, miR-125b-5p) and 7 protein-coding genes (TET1, TET2, TET3, TDG, DNMT3A, DNMT3B and DNMT1). The probability of their interaction has been estimated. Theoretical data have been obtained on the combination of triplets (axes) possibly involved in the development of ccRCC.

Funding: The study is supported by the FGFU-2022-0007 state assignment.

Список литературы/References

1. Клаг У.С., Каммингс М.Р., Спенсер Ш.А., Палладино М.А. Основы генетики. М.: Техносфера, 2021
   [Klug U.S., Cummings M.R., Spencer S.A., Palladino M.A. Fundamentals of Genetics. Moscow: Technosphere, 2021 (in Russian)]
2. Максимова В.П., Бугаева П.Е., Жидкова Е.М. и др. Современные подходы к выявлению и изучению эпигенетически активных ксенобиотиков. *Успехи молекулярной онкологии*. 2019;6(3):8-27
   [Maksimova V.P., Bugaeva P.E., Zhidkova E.M. et al. Modern approaches to the identification and study of epigenetically active xenobiotics. *Uspehi Molekulârnoj Onkologii = Successes of Molecular Oncology*. 2019;6(3):8-27 (in Russian)]
3. Tang Z., Li C., Kang B., Gao G., Li C., Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98-W102. doi 10.1093/nar/gkx247

doi 10.18699/bgrs2024-9.1-10

Активация термогенных футильных циклов в адипоцитах при гиперэкспрессии креатинкиназы B и глицеролкиназы

Мичурина С.[1]*, Стафеев Ю.[1], Белоглазова И.[1], Гольцева Ю.[1, 2], Дергилев К.[1], Парфёнова Е.[1, 2]

[1] *Национальный медицинский исследовательский центр кардиологии им. акад. Е.И. Чазова МЗ РФ, Москва, Россия*

[2] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

** michurinas192@gmail.ru*

Ключевые слова: термогенез; креатинкиназа B; глицеролкиназа; адипоциты

Мотивация и цель: Термогенез в жировой ткани представляет собой процесс диссипации избыточной энергии окисления глюкозы, жирных кислот и других субстратов в виде тепла. Благодаря анорексигенному и гипогликемическому эффекту термогенез является перспективной мишенью для разработки новых терапевтических подходов для борьбы с ожирением, сахарным диабетом 2 типа и сопутствующими метаболическими заболеваниями. Наряду с каноническим механизмом термогенеза, осуществляемым белком разобщителем внутренней мембраны митохондрий UCP1, продукция тепла в адипоцитах может увеличиваться при повышении активности футильных циклов [1–3]. Мы предположили, что активация термогенных футильных циклов, включающих креатинкиназный и триацилглицеридный циклы, может способствовать изменению углеводного и липидного метаболизма жировых клеток, усиливать утилизацию глюкозы и триглицеридов в окислительных процессах. Цель представленной работы заключается в активации экспрессии ключевых ферментов футильных циклов (креатинкиназы B (CKB) и глицеролкиназы (ГК)) в адипоцитах и анализе метаболизма модифицированных клеток [1, 4].

Методы и алгоритмы: Исследование проводили на культуре клеток 3T3-L1, дифференцированных в адипоциты под действием инсулина, изобутилметилксантина, дексаметазона и розиглитазона. Для получения лентивирусных конструкций для активации ГК и CKB белок-кодирующие последовательности мРНК NM_008194.3, NM_021273.4 были амплифицированы с помощью ПЦР с обратной транскрипцией и клонированы в вектор Lego-iG2, на основе которого были получены лентивирусные частицы. Зрелые адипоциты трансдуцировали и через 48 ч оценивали регуляцию экспрессии целевых белков методами ПЦР в реальном времени и иммуноблоттинга. В модифицированных клетках исследовали активность поглощения глюкозы, липолиза и липогенеза методом радиоизотопного анализа с применением [^{3}H]-2-дезоксиглюкозы и [^{14}C]-глюкозы. Метаболические параметры клеток оценивали в базальном состоянии и при стимуляции важнейшими регуляторами метаболизма адипоцитов: инсулином и норадреналином. Влияние гиперэкспрессии CKB и ГК на термогенез, мембранный потенциал митохондрий и накопление липидов оценивали методом конфокальной микроскопии при визуализации клеток с помощью флуоресцентных зондов ERthermAC, JC-1 и BODIPY493/503. Обработку

микрофотографий и количественный анализ проводили в программе Image ExFluorer с использованием машинного обучения.
Результаты: Полученные генетические конструкции увеличивают экспрессию СКВ и ГК в зрелых адипоцитах. Показано, что в клетках с гиперэкспрессией СКВ и ГК усиливается поглощение глюкозы в ответ на норадреналин, но не происходит изменение активности липолиза, липогенеза и инсулиновой чувствительности. Было продемонстрировано, что повышение экспрессии белков СКВ и ГК в адипоцитах ассоциировано с уменьшением количества мелких и средних липидных капель за счет увеличения экспрессии активатора слияния капель – белка FSP27. Гиперэкспрессия СКВ и ГК не оказывала влияние на базальный уровень термогенеза, но усиливала продукцию тепла при стимуляции норадреналином в сравнении с контрольными клетками. Кроме того, при активации экспрессии СКВ и ГК было обнаружено увеличение количества митохондрий с высоким мембранным потенциалом.
Выводы: При активации экспрессии СКВ и ГК, которые являются ключевыми ферментами креатинкиназного и триацилглицеридного футильных циклов, происходит повышение чувствительности клеток к норадреналину. В ответ на стимуляцию адренорецепторов в адипоцитах с повышенной экспрессией СКВ и ГК происходит более сильная активация термогенеза и поглощения глюкозы, чем в контрольных клетках. Повышение мембранного потенциала митохондрий в модифицированных клетках может быть связано с необходимостью в увеличении синтеза АТФ, так как активность футильных циклов сопряжена с расходованием энергии. Таким образом, активация экспрессии СКВ и ГК может быть использована для получения термогенных адипоцитов, характеризующихся высоким уровнем поглощения и утилизации глюкозы. На основе полученных клеток возможна разработка новых подходов в генной и клеточной терапии ожирения и сопутствующих метаболических заболеваний.
Финансирование: Исследование поддержано грантом РНФ № 22-75-10085.

Overexpression of creatine kinase B and glycerol kinase activates thermogenic futile cycles in adipocytes

Michurina S.[1]*, Stafeev I.[1], Beloglazova I.[1], Goltseva Y.[1, 2], Dergilev K.[1], Parfyonova Ye.[1, 2]

[1] *National Medical Research Centre of Cardiology named after academician E.I. Chazov, Moscow, Russia*
[2] *Lomonosov Moscow State University, Moscow, Russia*
* *michurinas192@gmail.ru*

Key words: thermogenesis; creatine kinase B; glycerol kinase; adipocytes

Motivation and Aim: Thermogenesis in adipose tissue dissipates excessive energy of glucose and fatty acids oxidation in the form of heat. Due to its anorexigenic and hypoglycemic effects, thermogenesis is a promising target for the development of new therapeutic approaches to combat obesity, type 2 diabetes mellitus and concomitant metabolic diseases. Along with the canonical mechanism of thermogenesis carried out by the uncoupling protein of mitochondrial inner membrane UCP1, heat production in adipocytes can be increased during futile cycles [1–3]. We suggest that activation of

thermogenic futile cycles, including creatine kinase and triacylglyceride cycles, may contribute to changes in carbohydrate and lipid metabolism of fat cells, enhance the utilization of glucose and triglycerides in oxidative processes. The purpose of the presented work is to activate the expression of key enzymes of the futile cycles (creatine kinase B (CKB) and glycerol kinase (Gyk)) in adipocytes and analyze the metabolism of modified cells [1, 4].

Methods and Algorithms: The study was performed in 3T3-L1 cells differentiated into adipocytes using insulin, isobutylmethylxanthine, dexamethasone and rosiglitazone. To obtain lentiviral constructs for Gyk and CKB overexpression, the protein-coding mRNA sequences NM_008194.3, NM_021273.4 were cloned into the Lego-iG2 vector using RT PCR. Mature adipocytes were transduced, and the expression of target proteins was measured in 48h by real-time PCR and immunoblotting. In modified cells glucose uptake, lipolysis and lipogenesis were analyzed using [^{3}H]-2-deoxyglucose and [^{14}C]-glucose. Metabolic parameters were measured in the basal state and under stimulation with important regulators of adipocyte metabolism: insulin and norepinephrine. Thermogenesis, mitochondrial membrane potential and lipid accumulation were examined by confocal microscopy of cells stained with ERthermAC, JC-1 and BODIPY493/503. Processing of microphotographs and quantification were carried out in Image ExFluorer using machine learning.

Results: The obtained genetic constructs increase the expression of CKB and Gyk in mature adipocytes. Adipocytes overexpressing CKB and Gyk has demonstrated enhanced glucose uptake in response to norepinephrine. However, they CKB and Gyk had no effect on lipolysis, lipogenesis and insulin sensitivity. Increased expression of CKB and Gyk in adipocytes was associated with a decrease in the number of small and medium sized lipid droplets due to increased expression of the droplet fusion activator FSP27. Overexpression of CKB and Gyk had no effect on the basal level of thermogenesis, but expression of both proteins increased heat production in response to norepinephrine. In addition, activation of CKB and Gyk expression increased the number of high membrane potential mitochondria in adipocytes.

Conclusion: Activation of CKB and Gyk, which are the key enzymes of the creatine kinase and triacylglyceride futile cycles, increases adipocytes norepinephrine sensitivity. Adipocytes overexpressing CKB and Gyk have higher level of thermogenesis and glucose uptake activation in response to adrenergic stimulation. An increase in the mitochondrial membrane potential in genetically modified cells could be associated with the increased ATP synthesis, since the activity of futile cycles is associated with elevated energy demand. Thus, activation of CKB and Gyk can be used to develop thermogenic adipocytes with high level of glucose consumption and utilization. Transplantation of obtained thermogenic adipocytes can become a new approache in gene and cell therapy of obesity and related metabolic disorders.

Funding: The study is supported by RSF grant No. 22-75-10085.

Список литературы/References

1. Kazak L. Promoting metabolic inefficiency for metabolic disease. *iScience*. 2023;26(10):107843. doi 10.1016/j.isci.2023.107843
2. Chouchani E.T., Kazak L., Spiegelman B.M. New advances in adaptive thermogenesis: UCP1 and beyond. *Cell Metabolism*. 2019;29(1):27-37
3. Bardova K., Funda J. et al. Additive effects of omega-3 fatty acids and thiazolidinediones in mice fed a high-fat diet: triacylglycerol/fatty acid cycling in adipose tissue. *Nutrients*. 2020;12(12):3737
4. Leroyer S.N. et al. Rosiglitazone controls fatty acid cycling in human adipose tissue by means of glyceroneogenesis and glycerol phosphorylation. *J Biol Chem*. 2006;281(19):13141

doi 10.18699/bgrs2024-9.1-11

Количественные различия в белковом спектре сыворотки крови больных БАР с депрессивным и смешанным эпизодом

Серегин А.А.[1]*, Смирнова Л.П.[1], Дмитриева Е.М.[1], Завьялова М.Г.[2], Рыжкова А.Ю.[3], Иванова С.А.[1]

[1] *Томский национальный исследовательский медицинский центр Российской академии наук, НИИ психического здоровья, Томск, Россия*

[2] *Сколковский институт науки и технологий, Москва, Россия*

[3] *Сибирский государственный медицинский университет Министерства здравоохранения Российской Федерации, Томск, Россия*

* *apocalips1991@mail.ru*

Ключевые слова: биполярное аффективное расстройство; протеом; масс-спектрометрия, биомаркеры, сыворотка крови

Мотивация и цель: Диагностика биполярного аффективного расстройства (БАР) является довольно трудной задачей так как БАР нередко имеет сходную клиническую картину с другими психическими расстройствами (ПР), а лабораторные критерии дифференциальной диагностики отсутствуют [1]. Симптоматика гипомании при униполярной депрессии со смешанными чертами имеет схожие клинические симптомы с гипоманией при БАР [2]. При лечении антидепрессантами предполагаемой униполярной депрессии со смешанными чертами есть большая вероятность ее трансформации в эпизод мании или гипомании при БАР [3]. Это связано с недостаточным пониманием механизмов патогенеза БАР и выявление регуляторных белков, участвующих в патогенезе БАР, в доступном для использования в диагностических целях биоматериале – сыворотке крови, не только приблизит к пониманию патогенетических механизмов БАР, но и поможет в разработке новых методов диагностики и патогенетически обоснованных методов терапии [4].

Методы и алгоритмы: В работе проанализировали образцы сыворотки крови 59 пациентов с БАР (13 мужчин и 46 женщин) в возрасте 37 [26; 51] лет и продолжительностью болезни 7 [3; 15] лет. Из данной группы 30 пациентов обратились в стационар по поводу текущего эпизода легкой или умеренной депрессии (F31.3), а 29 пациентов по поводу текущего эпизода смешанного характера (F31.6). Группу контроля составили 43 психически и соматически здоровых добровольцев, соответствующих полу и возрасту обследуемой группе. Забор крови, проводился до начала терапии. Из данной выборки масс-спектрометрическому анализу была подвергнута сыворотка крови 10 человек больных БАР и 5 здоровых лиц. Сыворотка при помощи аффинной хроматографии очищалась от 14 мажорных белков (колонка Multiple Affinity Removal Column Human 14, Agilent). Полученные образцы разделяли 1D электрофорезом в 12 % полиакриламидном геле по методу Леммли. Далее после трипсинолиза и экстракции пептидов из геля были получены спектры масс белков при помощи ВЭЖХ/масс-спектрометрии с использованием системы ВЭЖХ Ultimate 3000 Nano

LC (Thermo Scientific), соединенной с масс-спектрометром Q Exactive HF-X – Orbitrap (Thermo Fischer Scientific) на базе ЦКП передовой масс-спектрометрии Центра исследовательской инфраструктуры Сколковского института науки и технологий, г. Москва. Идентификация белков проводилась с использованием базы данных UniProtKB/Swiss-Prot (www.uniprot.org) и поисковой машины Mascot (www.matrixscience.com). Методом твердофазного иммуноферментного анализа (ИФА) в образцах сыворотки крови больных БАР и здоровых лиц было определено количество следующих белков: альфа-актин-2 (a-SMA, a-Smooth Muscle Actin, E01S0004 Cloud-Clone Corp., США); дермцидин (DCD, Dermcidin, SEC896Hu Cloud-Clone Corp., США); белок родственный рецептору липопротеинов низкой плотности/мегалин (Low Density Lipoprotein Receptor Related Protein 2 (LRP2, WED101Hu Cloud-Clone Corp., USA). Количество перечисленных белков измеряли согласно протоколу производителя наборов ИФА на многорежимном ридере Varioskan LUX (Thermo Scientific, США) на базе ЦКП «Медицинская геномика» (ТНИМЦ). Статистическую значимость результата ИФА проверяли с помощью непараметрического анализа ANOVA (критерий Краскела-Уоллиса) для трех независимых групп и критерия Манна-Уитни для двух независимых групп.

Результаты: В результате масс-спектрометрического анализа было выявлено около 1600 белков в каждой исследуемой группе. При сопоставлении протеомов у пациентов с БАР был идентифицирован 21 дифференциально экспрессируемый патогенетически значимый белок. Данные белки являются структурными элементами цитоскелета и мембранных рецепторов, участвуют в регуляции синтеза ДНК и клеточного цикла, дифференцировке нервных клеток и процессах транспорта через клеточную мембрану. Для нескольких из выявленных белков уже установлено участие в патогенезе психических расстройств. Для дальнейшего количественного анализа были выбраны три наиболее интересных белка: a-SMA, DCD и LRP2. При помощи статистического анализа ANOVA (критерий Краскела–Уоллиса) были обнаружены статистически значимые различия между подгруппами больных БАР с текущем депрессивным и смешанным эпизодом, а также со здоровыми лицами в содержании белков LRP2 (p = 0,016) и DCD (p = 0,049). При попарном сравнении исследуемых подгрупп при помощи критерия Манна–Уитни выявлено что разница в содержании этих белков возникает за счет повышения уровня LRP2 (p = 0,006) и DCD (p = 0,027) у больных БАР с текущем депрессивным эпизодом в сравнении со здоровыми лицами. Так дермцидин (DCD) активирует репарацию нейронов в случае их повреждения [4]. Известно, что прекурсор дермцидина – пептид Y-P30 [5] оказывает нейропротекторное действие [6] и способствует росту нейритов из нейронов-предшественников таламуса и мозжечка [7]. LRP2 в свою очередь регулирует развитие и миграцию нейрональных стволовых клеток, расположенных в двух ключевых нишах взрослого организма: субвентрикулярной зоне боковых желудочков и зубчатой извилине гиппокампа, а также может играть регенеративную и проективную роль при повреждении нейронов [9]. Несмотря на то, что анализ ANOVA не выявил статистически значимых различий в содержании альфа-актина-2 в исследуемых подгруппах, при попарном сравнении выявлено значительное повышение этого белка у больных БАР с текущим смешанным эпизодом в сравнении со здоровыми лицами (Mann–Whitney U Test, p = 0,0422). Существует предположение о возможном участии этого белка в регуляции проводимости NMDA-рецепторов

[10]. Также известно, что лечение литием снижает уровни экспрессии α-SMA [11]. По анамнестическим данным включенные в эксперимент пациенты не принимали лечение как минимум на протяжении двух месяцев до поступления.
Выводы: Обнаруженное в данной работе повышенное содержание LRP2 и DCD у больных БАР с текущем депрессивным эпизодом может указывать на запуск защитных нейропротективых механизмов в ответ на повреждение нейронов в процессе патогенеза БАР, однако нейропротективная роль данных белков в патогенезе БАР требует детального изучения. Нейропротективное действие DCD а также возможное участие α-SMA в регуляции проводимости NMDA-рецепторов позволяет предложить данные белки в качестве потенциального биомаркера БАР при более подробном изучении их патогенетической роли при БАР.
Финансирование: Исследование поддержано грантом РНФ № 23-75-00023.

Quantitative differences in the protein spectrum of blood serum of patients with bipolar disorder with depressive and mixed episodes

Seregin A.A.[1]*, Smirnova L.P.[1], Dmitrieva E.M.[1], Zavyalova M.G.[2], Ryzhkova A.Yu.[3], Ivanova S.A.[1]

[1] *Tomsk National Research Medical Canter, Russian Academy of Sciences, Mental Health Research Institute, Tomsk, Russia*

[2] *Skolkovo Institute of Science and Technology, Moscow, Russia*

[3] *Siberian State Medical University, SSMU, Tomsk, Russia*

* *ivanov@bionet.nsc.ru*

Key words: bipolar disorder; proteome; mass spectrometry, biomarkers, blood serum

Motivation and Aim: Diagnosis of bipolar disorder (BD) is a rather difficult task since BD often has a similar clinical picture with other mental disorders (MD), and laboratory criteria for differential diagnosis are lacking [1]. The symptoms of hypomania in unipolar depression with mixed features have similar clinical symptoms to hypomania in bipolar disorder [2]. When treated with antidepressants for suspected unipolar depression with mixed features, there is a high probability of its transformation into an episode of mania or hypomania in BD [3]. This is due to insufficient understanding of the mechanisms of BD pathogenesis. Identification of regulatory proteins involved in the pathogenesis of BD in biomaterial available for diagnostic purposes – blood serum – will not only bring us closer to understanding the pathogenetic mechanisms of BD, but will also help in the development of new pathogenetically based methods of diagnosis and therapy [4].
Methods and Algorithms: The study analyzed blood serum samples from 59 patients with BD (13 men and 46 women) aged 37 [26; 51] years and disease duration 7 [3; 15] years. Of this group, 30 patients visited the hospital for a current episode of mild or moderate depression (F31.3), and 29 patients for a current mixed episode (F31.6). The control group consisted of 43 mentally and somatically healthy volunteers corresponding to the gender and age of the study group. Blood sampling was carried out before the start of therapy. From this sample, the blood serum of 10 people with BD and 5 healthy individuals was subjected to mass spectrometric analysis. The serum was purified from

14 major proteins using affinity chromatography (Multiple Affinity Removal Column Human 14, Agilent). The resulting samples were separated by 1D electrophoresis in a 12 % polyacrylamide gel according to the Laemmli method. Next, after trypsinolysis and extraction of peptides from the gel, protein mass spectra were obtained by HPLC/mass spectrometry using an Ultimate 3000 Nano LC HPLC system (Thermo Scientific) coupled to a Q Exactive HF-X – Orbitrap mass spectrometer (Thermo Fischer Scientific) based on the Advanced Mass Spectrometry Core Facility of Skolkovo Institute of Science and Technology, Moscow. Protein identification was performed using the UniProtKB/Swiss-Prot database (www.uniprot.org) and the Mascot search engine (www.matrixscience.com). Using enzyme-linked immunosorbent assay (ELISA), the levels of the following proteins were determined in blood serum samples of patients with bipolar disorder and healthy individuals: Actin, aortic smooth muscle (a-SMA, a-Smooth Muscle Actin, E01S0004 Cloud-Clone Corp., USA); dermcidin (DCD, Dermcidin, SEC896Hu Cloud-Clone Corp., USA); Low-density lipoprotein receptor-related protein 2 (LRP2) WED101Hu Cloud-Clone Corp., USA). The level of the listed proteins was measured according to the protocol of the manufacturer of ELISA kits on a multimode reader Varioskan LUX (Thermo Scientific, USA) on the basis of the Medical Genomics Center (Tomsk National Research Medical Center of the Russian Academy of Sciences). The statistical significance of the ELISA result was tested using nonparametric ANOVA analysis (Kruskal-Wallis test) for three independent groups and the Mann-Whitney test for two independent groups.

Results: As a result of mass spectrometric analysis, about 1600 proteins were identified in each study group. Comparison of proteomes in patients with BD identified 21 differentially expressed pathogenetically significant proteins. These proteins are structural elements of the cytoskeleton and membrane receptors; they are involved in the regulation of DNA synthesis and the cell cycle, differentiation of nerve cells and transport processes across the cell membrane. Several identified proteins have already been shown to be involved in the pathogenesis of mental disorders. The three most interesting proteins were selected for further quantitative analysis: a-SMA, DCD, and LRP2. Using ANOVA statistical analysis (Kruskal–Wallis test), statistically significant differences were found between the subgroups of patients with bipolar disorder with a current depressive and mixed episode, as well as with healthy individuals in the content of LRP2 ($p = 0.016$) and DCD ($p = 0.049$) proteins. When pairwise comparison of the studied subgroups using the Mann–Whitney test, it was revealed that the difference in the content of these proteins arises due to an increase in the level of LRP2 ($p = 0.006$) and DCD ($p = 0.027$) in patients with bipolar disorder with a current depressive episode in comparison with healthy individuals. Thus, dermcidin (DCD) activates the repair of neurons in case of their damage [4]. It is known that the precursor of dermcidin, the Y-P30 peptide [5], has a neuroprotective effect [6] and promotes the growth of neurites from the precursor neurons of the thalamus and cerebellum [7]. LRP2, in turn, regulates the development and migration of neuronal stem cells located in two key niches of the adult body: the subventricular zone of the lateral ventricles and the dentate gyrus of the hippocampus, and can also play a regenerative and projective role in neuronal damage [9]. Even though ANOVA analysis did not reveal statistically significant differences in the content of α-SMA in the studied subgroups, pairwise comparison revealed a significant increase in this protein in patients with bipolar disorder with a current mixed episode compared with healthy individuals (Mann–Whitney U Test, $p = 0.0422$). There is an assumption about the possible participation of this protein in the regulation of the

conductivity of NMDA receptors [10]. Lithium treatment is also known to reduce α-SMA expression levels [11]. According to anamnestic data, the patients included in the experiment did not take treatment for at least 2 months before admission.

Conclusion: The increased content of LRP2 and DCD found in this work in BD patients with a current depressive episode may indicate the launch of protective neuroprotective mechanisms in response to neuronal damage during the pathogenesis of BD, however, the neuroprotective role of these proteins in the pathogenesis of BD requires detailed study. The neuroprotective effect of DCD as well as the possible involvement of α-SMA in the regulation of NMDA receptor conductance allows us to propose these proteins as a potential biomarker of BD with a more detailed study of their pathogenetic role in BD.

Funding: The study is supported by Grant of RSF No. 23-75-00023.

Список литературы/References

1. Geoffroy P.A., Leboyer M., Scott J. Predicting bipolar disorder: what can we learn from prospective cohort studies? *Encephale*. 2015;41:10-16. doi 10.1016/j.encep.2013.05.004
2. Benazzi F. Mixed depression: a clinical marker of bipolar II disorder. *Prog Neuropsychopharmacol. Biol Psychiatry*. 2005;29:267-274
3. Bottlernder R. et al. Mixed depressive features predict maniform switch during treatment of depression in bipolar I disorder. *J Affect Disord*. 2004;78:149-152
4. Lee Motoyama J.P. et al. Identification of dermcidin in human gestational tissue and characterization of its proteolytic activity. *Biochem Biophys Res Commun*. 2007;357:828-833
5. Cunningham T.J. et al. Identification of the human cDNA for new survival/evasion peptide (DSEP): studies *in vitro* and *in vivo* of overexpression by neural cells. *Exp Neurol*. 2002;177:32-39
6. Macharadze T. et al. Y-P30 confers neuroprotection after optic nerve crush in adult rats. *Neuroreport*. 2011;22:544-547
7. Landgraf P. et al. The survival-promoting peptide Y-P30 enhances binding of pleiotrophin to syndecan-2 and -3 and supports its neuritogenic activity. *J Biol Chem*. 2008;283:25036-25045
8. Pollen A.A. et al. Molecular identity of human outer radial glia during cortical development. *Cell*. 2015;1(163):55-67
9. O'Rourke M. et al. Adult myelination: wrapping up neuronal plasticity. *Neural Regen Res*. 2014;9(13):1261-1264. doi 10.4103/1673-5374.137571
10. Логинова Л.В. Связь белков сыворотки крови, выявленных с помощью протеомного анализа, с особенностями патогенеза психогенных и эндогенных психических расстройств: дис. ... канд. мед. наук. Томск, 2018
11. Chen P.-H., Chung C.-C., Liu S.-H., Kao Y.-H., Chen Y.-J. Lithium treatment improves cardiac dysfunction in rats deprived of rapid eye movement sleep. *Int J Mol Sci*. 2022;23:11226. doi 10.3390/ijms231911226

doi 10.18699/bgrs2024-9.1-12

Протеомное профилирование нативных клапанов сердца, пораженных инфекционным эндокардитом

Синицкая А.В.*, Костюнин А.Е., Асанов М.А., Поддубняк А.О., Хуторная М.В., Синицкий М.Ю.

Научно-исследовательский институт комплексных проблем сердечно-сосудистых заболеваний, Кемерово, Россия

** annacepokina@mail.ru*

Ключевые слова: инфекционный эндокардит; масс-спектрометрия; протеом; дифференциально экспрессируемые белки; нативные клапаны сердца

Мотивация и цель: Инфекционный эндокардит (ИЭ) – воспалительное заболевание эндокарда клапанных структур сердца инфекционной природы, характеризующееся высоким риском осложнений. Несмотря на достижения в диагностике ИЭ, основанные на использовании современных гистологических и молекулярно-генетических методов, фундаментальные аспекты патогенеза этого заболевания изучены недостаточно хорошо.

Цель исследования: идентификация ключевых белков и сигнальных путей, вовлеченных в формирование ИЭ.

Методы и алгоритмы: Материалом для исследования послужили иссеченные в ходе кардиохирургической операции по протезированию аортальные клапаны (АК), пораженные ИЭ (n = 3) или кальцинирующим аортальным стенозом (КАС) (n = 3). Выделение белка из фрагментов клапанов сердца проводили путем их гомогенизации в лизирующем растворе T-PER с добавлением коктейля ингибиторов протеиназ и фосфотаз Halt. Количество выделенного белка оценивали с использованием набора BCA Protein Assay Kit. Протеомное профилирование проводили с использованием высокоэффективной жидкостной хроматографии в сочетании с масс-спектрометрией (ВЭЖХ-МС) на масс-спектрофотометре TimsToF Pro. Идентификацию белков осуществляли с использованием программного обеспечения Peaks Xpro v.10.6. Достоверными считали идентификации белков с FDR <1 % и наличием как минимум двух уникальных пептидов. Для анализа использовали базы данных SwissProt и cRAP, отфильтрованные по белкам человека. После получения информации о площади пиков AUC, обнаруженных в проанализированных образцах белков, проводили количественный анализ полученных данных в программной среде R v.4.1.2. Белки с более чем 2/3 пропущенными значениями были исключены из анализа. Анализ дифференциальной экспрессии проводили с использованием пакета «limma» (v.3.50.3). Анализ главных компонент (PCA) проводился с использованием пакета «mixOmics» (v.6.18.1). Для визуализации данных использовались пакеты «ggplot2» (v.3.4.4) и «EnhancedVolcano» (v.1.12.0). Анализ обогащения путей дифференциально-экспрессируемых белков (DEPs) проводился с использованием анализа генной онтологии (https://geneontology.org/, по состоянию на 29 апреля 2024 г.) и базы данных Reactome (https://reactome.org/, по состоянию на 29 апреля 2024 г.).

Результаты: В тканях нативных клапанов сердца, пораженных ИЭ, при сравнительном анализе с клапанами, пораженными КАС, выявлено 272 DEPs, удовлетворяющих следующим критериям: log2-кратное изменение экспрессии >1 и значение p <0.05, скорректированное с помощью поправки FDR, из них 53.68 % DEPs были гипоэкспрессированы, 46.32 % – гиперэкспрессированы в тканях клапанов с ИЭ. Для 23 DEPs (PLOD, FNDC1, ADRM1, ACDSB, HARS1, NEB2, TPC6B, RBM4, EMIL1, DNJA2, SGTB, H32, RU2B, RS15A, RL15, ACTS, BT3L4, SCPDL, TRXR2, NASP, PKHO2, STMN1 и P4HA2) показано увеличение экспрессии более чем в 5 раз, а для 30 DEPs (FTO, HV321, LV39, CLC11, VP13C, HV333, PGM5, SUSD2, ACTA, CAH3, LV469, HV353, SPT6H, AL1A2, MAMC2, ASPN, HBS1L, IBP6, MINP1, CK068, NENF, PCOC2, FABP4, PLS3, SCRN2, TGFI1, CCD43, SPP24, BMPER и IF4E3) отмечено пятикратное снижение экспрессии относительно клапанов с КАС. Аннотирование выявленных DEPs с использованием базы Reactome и анализа обогащения (Gene Ontology Enrichment Analysis) показало, что идентифицированные DEPs задействованы в следующих биологических процессах: регуляция системы комплемента, организация внеклеточного матрикса, образование коллагена, инициация врожденного иммунного ответа, активация системы комплемента, везикулярный транспорт, клеточный ответ на паразитарные инфекции, передача регуляторных сигналов, опосредуемых В-клеточным рецептором.
Выводы: Установлены ключевые белки и сигнальные пути, вовлеченные в формирование дисфункций клапанного аппарата сердца, вызванных различным этиологическим агентами.
Финансирование: Исследование выполнено на средства гранта Российского научного фонда № 23-75-10020 «Молекулярно-генетические основы патогенеза инфекционного эндокардита нативных клапанов сердца и их биопротезов», https://rscf.ru/project/23-75-10020/.

Proteomic profiling of native heart valves affected by infective endocarditis

Sinitskaya A.*, Kostyunin A., Asanov M., Poddubnyak A., Khutornaya M., Sinitsky M.

Research Institute for Complex Issues of Cardiovascular Diseases, Kemerovo, Russia
** annacepokina@mail.ru*

Key words: infective endocarditis; mass spectrometry; proteome; differentially expressed proteins; native heart valves

Motivation and Aim: Infective endocarditis (IE) is an inflammatory disease caused by infections and affecting the endocardium of heart valves. IE is characterized by a high risk of complications. Despite advances in the IE diagnosis based on the modern histological and molecular genetic methods, the fundamental aspects of its pathogenesis are still not fully understood. The presented study *was aimed* to the identification of key proteins and signaling pathways involved in IE pathogenesis.
Methods and Algorithms: Aortic valves (AVs) explanted during cardiac surgery due to IE (n = 3) or calcific aortic valve disease (CAVD) (n = 3) were used in the presented

study. To protein isolation, the fragments of heart valves were homogenized in T-PER lysing buffer supplied with a Halt protease and phosphatase inhibitor cocktail. The quantity of isolated protein was evaluated by a BCA Protein Assay Kit. Proteomic profiling was performed by ultra-high performance liquid chromatography-mass spectrometry using a TimsToF Pro mass spectrometer. Protein identification was performed using Peaks Xpro software v.10.6. Proteins characterized by FDR <1 % and presence at least two unique peptides were selected for further analysis. For data analysis, we used human protein SwissProt database and cRAP contaminants database. Quantitative analysis of the obtained data was performed in R v.4.1.2 after estimation of areas under curves. Proteins with more than 2/3 missing values were excluded from the analysis. Differential expression analysis was performed using the «limma» package (v.3.50.3). Principal component analysis (PCA) was performed using the «mixOmics» package (v.6.18.1). «ggplot2» (v.3.4.4) and «EnhancedVolcano» (v.1.12.0) packages were used for data visualization. Pathway enrichment analysis of differentially expressed proteins (DEPs) was performed using Gene Ontology (https://geneontology.org/, accessed on 29 April 2024) and Reactome (https://reactome.org/, accessed on 29 April 2024) databases.

Results: In the tissues of native heart valves affected by IE, when it was compared to heart valves affected by CAVD, 272 DEPs were identified according to the following criteria: log2-fold change in expression >1 and p value <0.05, corrected by correction FDR. 53.68 % identified DEPs were downexpressed and 46.32 % were overexpressed in the heart valves with IE. For 23 DEPs (PLOD, FNDC1, ADRM1, ACDSB, HARS1, NEB2, TPC6B, RBM4, EMIL1, DNJA2, SGTB, H32, RU2B, RS15A, RL15, ACTS, BT3L4, SCPDL, TRXR2, NASP, PKHO2, STMN1 and P4HA2) more then 5-fold increased expression was shown, and for 30 DEPs (FTO, HV321, LV39, CLC11, VP13C, HV333, PGM5, SUSD2, ACTA, CAH3, LV469, HV353, SPT6H, AL1A2, MAMC2, ASPN, HBS1L, IBP6, MINP1, CK068, NENF, PCOC2, FABP4, PLS3, SCRN2, TGFI1, CCD43, SPP24, BMPER and IF4E3) more then 5-fold decreased expression compared to the heart valve with CAVD was shown. Annotation of the identified DEPs using the Reactome database and enrichment analysis (Gene Ontology Enrichment Analysis) showed that the identified DEPs are involved in the following biological processes: regulation of the complement system, organization of the extracellular matrix, collagen formation, initiation of the innate immune response, activation of the complement system, vesicular transport, cellular response to parasitic infections, transmission of regulatory signals mediated by the B-cell receptor.

Conclusion: The key proteins and signaling pathways involved in the pathogenesis of cardiac valvular dysfunction caused by various etiological agents have been identified.

Funding: The study is supported by the Grant of Russian Science Foundation No. 23-75-10020 “Molecular genetic basis of the pathogenesis of infective endocarditis in native and bioprosthetic heart valves”, https://rscf.ru/ project/23-75-10020/.

doi 10.18699/bgrs2024-9.1-13

Преэклампсия: взгляд на патогенез заболевания через призму геномики и транскриптомики

Трифонова Е.А.*, Бабовская А.А., Зарубин А.А., Сваровская М.Г., Степанов В.А.
Томский национальный исследовательский медицинский центр Российской академии наук (Томский НИМЦ), Томск, Россия
** ekaterina.trifonova@medgenetics.ru*

Ключевые слова: транскриптом; геном; преэклампсия; RNA-seq; tagSNPs

Мотивация и цель: Анализ генома и транскриптома различных тканей и клеток, а также регуляторных механизмов экспрессии генов представляет значительный интерес для исследователей, поскольку нарушение транскрипционной активности генов вовлечено в патогенез многочисленных заболеваний. На сегодняшний день омиксные технологии нашли широкое применение во всевозможных областях предиктивной медицины, к примеру, в разработке методов диагностики, стратегий прогнозирования и профилактики различных заболеваний, включая и акушерскую патологию, где в качестве одного из основных объектов изучения выступает плацента. Необходимо отметить, что изучение молекулярных процессов, происходящих в плацентарной ткани и связанных с транскрипционной регуляцией генной экспрессии, рассматривается на сегодняшний день как наиболее перспективное направление для раскрытия патофизиологии осложненного течения беременности, и в частности такого тяжелого заболевания как преэклампсия (ПЭ). В связи с чем целью нашей работы было охарактеризовать паттерны транскриптома плацентарной ткани и ее отдельных субпопуляций клеток, специфичные для женщин с ПЭ и физиологическим течением беременности и оценить роль полиморфных маркеров генов, дифференциально экспрессирующихся при патологичекой и нормальной беременности, в формировании структуры предрасположенности к ПЭ.
Методы и алгоритмы: Полногеномный анализ экспрессионных профилей 24 образцов дистальной части плацентарной ткани пациенток с ПЭ и женщин с физиологической беременностью (ФБ) выполнен с помощью технологии микрочипов (HT-12 BeadChip v.4, Illumina). Единичные децидуальные клетки (ДК) и клетки синцитиотрофобласта (СТБ) были получены с помощью технологии лазерной микродиссекциии («Laser Capture Microdissection») препаратов тонких окрашенных срезов. Тотальная РНК 1000 изолированных ДК и клеток СТБ была выделена набором Single Cell RNA Purification Kit (Norgen). Синтез библиотек для полнотранскриптомного анализа проводился набором SMARTer Stranded Total RNA-Seq Kit v2 (Takara). Готовые библиотеки объединяли в пул и секвенировали на платформе Illumina Next-seq 500. Для поиска дифференциально экспрессирующихся генов (ДЭГ) был использован метод обобщенных линейных моделей (GLM) программного пакета limma. Функциональный анализ кластера ДЭГ, проводили с помощью веб-инструмента Molecular Signatures Database (MSigDB). Конструирование генных сетей проводили с использованием программы STRING 9.0.

Суммарный объем выборки для анализа вариабельности ДНК, включенной в исследование, составил более 1000 женщин, которые согласно течению и исходам беременности были разделены на группу больных с ПЭ (N = 721 чел.) и контрольную группу (N = 552 чел.). Генотипирование полиморфных маркеров (tagSNPs) ДЭГ осуществляли с помощью MALDI-TOF масс-спектрометрии. Для сравнения частот аллелей и генотипов между анализируемыми группами использовали критерий χ^2 Пирсона с поправкой Йейтса или двусторонний точный тест Фишера.

Результаты: В плацентарной ткани обнаружено 63 гена, статистически значимо дифференциально экспрессирующихся между пациентками с ПЭ и физиологической беременностью. Функциональная аннотация ДЭГ свидетельствует о ряде биологических процессов, играющих важную роль в молекулярном патогенезе ПЭ: реакции, связанные с иммунным ответом, межклеточным взаимодействием, регуляция апоптоза и др. Анализ метаболических путей, в которые включены ДЭГ, указывает на возможное участие в патофизиологических механизмах ПЭ путей цитотоксичности, обусловленной NK-клетками, трансэндотелиальной миграции лейкоцитов и сигнальных путей, опосредованных активаторами GTP-аз. Полученная с помощью базы данных «STRING» сеть белок-белковых взаимодействий подтверждает взаимосвязи между генами, выявленные при исследовании биологических путей и процессов. Интеграция данных функциональной аннотации ДЭГ и анализа сетевых взаимодействий белков, кодируемых этими генами, позволила выделить кластер наиболее значимых ДЭГ, наследственная вариабельность которых была изучена на геномном уровне.

В частности, сравнительный анализ распределения частот аллелей и генотипов 85 полиморфных маркеров (tagSNPs) данных ДЭГ между пациентками с ПЭ и женщинами с ФБ выявил ассоциацию с развитием ПЭ 27 полиморфных вариантов следующих генов: *ANKRD37, BHLHE40, CORO2A, GPT2, HK2, INHA, LEP, LHB, NDRG1, PLIN2, PPP1R12C, SASH1, SIGLEC6, SYDE1* и *ZNF175*. Необходимо отметить, что большинство ассоциированных с ПЭ tagSNPs согласно результатам анализа с использованием онлайн ресурсов HaploReg и RegulomeDB были локализованы в сайтах связывания транскрипционных факторов и имели высокий регуляторный потенциал. Данное наблюдение представляет особый интерес в контексте результатов, полученных в ряде недавних работ, демонстрирующих важную роль регуляторных участков плацентарных генов как в развитии осложненного течения беременности, так и в отношении предрасположенности к множественным постнатальным заболеваниям человека. Тем не менее стоит отметить, что, исходя из современных представлений о крайне гетерогенном клеточном составе плаценты высоко актуальны исследования по картированию генетических вариантов, определяющих экспрессию генов в различных типах клеток плаценты при нормальной и патологической беременности.

Так, в результате проведенного нами анализа транскриптома на уровне отдельных субпопуляций клеток плацентарной ткани выявлены 26 генов, экспрессия которых статистически значимо различается в клетках СТБ женщин с ПЭ и физиологическим течением беременности, и 35 ДЭГ обнаружено для ДК. Кластер ДЭГ содержит не только известные гены-кандидаты, выявленные ранее во многих зарубежных полногеномных исследованиях плаценты (к примеру, *LEP, INHBA* и *FLT1*), но и новые гены (*AC098613.1, AC087857.1, FCRLB, TENM4, PTP4A1P7,*

LINC01225 и др.), которые могут рассматриваться в качестве новых биологических маркеров ПЭ и представляют интерес для дальнейшего изучения. Результаты функциональной аннотации ДЭГ показывают, что с развитием ПЭ на уровне СТБ могут быть связаны сигнальные пути регуляции гормональной секреции, MAPK-каскада, ERK1 и ERK2 каскада, положительной регуляция клеточной адгезии и пролиферации эндотелиальных клеток. Наряду с этим биологические процессы, задействованные в молекулярный патогенез ПЭ на уровне ДК, связаны с синтезом гамма-аминомасляной кислоты и развитием оксидативного стресса.
Выводы: Таким образом, в представленной работе обнаружены новые генетические маркеры ПЭ и показана значимая роль регуляторных участков генома человека в подверженности к данной патологии. На примере ПЭ апробирован системный подход к поиску перспективных биомаркеров гестационных осложнений, основанный на комбинации геномных, транскриптомных и биоинформатических методов. Это исследование расширяет представление о задействованных в ПЭ молекулярных механизмах и может служить основой для разработки профилактических, прогностических и терапевтических стратегий в области персонифицированного акушерства.
Финансирование: Исследование выполнено за счет средств государственного задания по теме ФНИ № 122020200083-8.

Preeclampsia: a view on the pathogenesis of disease through the prism of genomics and transcriptomics

Trifonova E.A.*, Babovskaya A.A., Zarubin A.A., Svarovskaya M.G., Stepanov V.A.
Research Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, Russia
** Ekaterina.trifonova@medgenetics.ru*

Key words: transcriptome; genome; preeclampsia; RNA-seq; tagSNPs

Motivation and Aim: The significant interest to researchers is the analysis of the genome and transcriptome of various tissues and cells as well as the regulatory mechanisms of gene expression since disruption of the transcriptional activity of genes is involved in the pathogenesis of numerous diseases. Today, omics technologies have found wide application in various areas of predictive medicine. For example, in the development of diagnostic methods, strategies for predicting and preventing various diseases including obstetric pathology where the placenta is one of the main objects of study. It should be noted that the study of molecular processes occurring in placental tissue and associated with the transcriptional regulation of gene expression is considered today as the most promising direction for revealing the pathophysiology of complicated pregnancy, and in particular such a serious disease as preeclampsia (PE). In this regard, the goal of our work was to characterize the transcriptome patterns of placental tissue and its individual cell subpopulations specific to women with PE and the physiological course of pregnancy and to evaluate the role of polymorphic gene markers differentially expressed in pathological and normal pregnancy in the formation of the structure of predisposition to PE.
Methods and Algorithms: Genome-wide analysis the expression profiles of 24 samples the distal part of the placental tissue from patients with PE and women with normal

pregnancy (NP) was performed using microarray technology (HT-12 BeadChip v.4, Illumina). Single decidual cells (DCs) and syncytiotrophoblast cells (SCTs) were obtained using laser microdissection technology ("Laser Capture Microdissection") of thin stained sections. Total RNA of 1000 isolated DCs and SCTs cells were extracted using the Single Cell RNA Purification Kit (Norgen). For full transcriptome analysis the library synthesis was carried out using the SMARTer Stranded Total RNA-Seq Kit v2 (Takara). The libraries were pooled and sequenced on the Illumina Next-seq 500 platform. The generalized linear models (GLM) method of the limma software package was used to search for differentially expressed genes (DEGs). Functional analysis of the DEG cluster was performed using the Molecular Signatures Database (MSigDB) web tool. Gene networks were constructed using the STRING 9.0 program.

The total sample size for the analysis of DNA variability included in the study was more than 1000 women were divided into a group of patients with PE (N = 721 people) and a control group (N = 552 people) according the outcomes of pregnancy. Genotyping of polymorphic markers (tagSNPs) of DEGs was carried out using MALDI-TOF mass spectrometry. To compare allele and genotype frequencies between the analyzed groups were used Pearson's χ^2 test with Yates' correction or Fisher's exact test.

Results: We found 63 genes that were statistically significantly differentially expressed between patients with PE and physiological pregnancy in placental tissue.

The functional annotation of DEGs indicates a number of biological processes that play an important role in the molecular pathogenesis of PE: reactions associated with the immune response, intercellular interaction, regulation of apoptosis, etc. Analysis of metabolic pathways in which DEGs are included indicates a possible participation in the pathophysiological mechanisms of PE pathways cytotoxicity by NK cells, transendothelial migration of leukocytes and signaling pathways mediated by activators of GTPases. The network of protein-protein interactions obtained using the STRING database confirms the relationships between genes identified in the study of biological pathways and processes. Integration the functional annotation data and analysis of proteins interactions encoded made it possible to identify a cluster of the most significant DEGs with the hereditary variability at the genomic level.

Thus, a comparative analysis of the distribution frequencies of alleles and genotypes of 85 polymorphic markers (tagSNPs) DEGs between patients with PE and NP revealed an association with the PE of 27 polymorphic variants of the following genes: *ANKRD37, BHLHE40, CORO2A, GPT2, HK2, INHA, LEP, LHB, NDRG1, PLIN2, PPP1R12C, SASH1, SIGLEC6, SYDE1* and *ZNF175*. It should be noted that the majority of associated tagSNPs were localized in transcription factor binding sites and had a high regulatory potential according of analysis using the online resources HaploReg and RegulomeDB. However, studies on mapping genetic variants that determine gene expression in various types of placental cells during normal and pathological pregnancy are highly relevant based on modern ideas about the extremely heterogeneous cellular composition of the placenta.

As a result of our analysis of the transcriptome of individual subpopulations cells of placental tissue 26 genes were identified. Expression these genes is statistically significantly different in the SCTs of women with PE and the NP, and 35 DEGs were found for DC. The DEG cluster contains not only well-known candidate genes previously identified in many foreign genome-wide studies of the placenta (for example, *LEP, INHBA* and *FLT1*) but also new genes (*AC098613.1, AC087857.1, FCRLB, TENM4, PTP4A1P7, LINC01225* and etc.), which can be considered as new biological

markers of PE and interesting for further study. The results of functional annotation of DEGs show that the development of PE at the level of SCTs may be associated with signaling pathways regulating hormonal secretion, MAPK cascade, ERK1 and ERK2 cascade, positive regulation of cell adhesion and proliferation of endothelial cells. Along with this, the biological processes involved in the molecular pathogenesis of PE at the DCs level are associated with the synthesis of gamma-aminobutyric acid and the development of oxidative stress.

Conclusion: Thus, in the presented work were discovered new genetic markers of PE and was shown the significant role of regulatory regions of the human genome in susceptibility to this pathology. Using the example of PE was tested a systematic approach to the search for promising biomarkers of gestational complications witch based on a combination of genomic, transcriptomic and bioinformatics methods. This study expands our understanding of the molecular mechanisms involved in PE and may serve as a basis for the development of preventive, prognostic and therapeutic strategies in the personalized obstetrics.

Funding: The research was carried out at the expense of the state assignment (fundamental scientific research No. 122020200083-8).

doi 10.18699/bgrs2024-9.1-14

Поиск генов-мишеней микроРНК во взаимодействии человека и кишечного микробиома

Трошина Д.А.[2]*, Шошин Ф.В.[3], Степанова А.А.[1], Орлов Ю.Л.[1, 4]

[1] *Первый МГМУ им. И.М. Сеченова Минздрава России, Москва, Россия*

[2] *Московский государственный университет имени М.В. Ломоносова, Москва, Россия*

[3] *ФГБОУ ВО ЯГМУ Минздрава России, Ярославль, Россия*

[4] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

* *darya.troshina02@mail.ru, orlov@bionet.nsc.ru*

Ключевые слова: медицинская информатика; гены-мишени микроРНК; кишечный микробиом; микроРНК; IntaRNA; TargetRNA3

Мотивация и цель: Современная медицина сталкивается с необходимостью разработки новых подходов к профилактике и лечению широкого спектра заболеваний, связанных с дисбалансом кишечного микробиома. МикроРНК, как ключевые регуляторы генной экспрессии [1], представляют собой перспективные молекулярные инструменты для модулирования микробиоты [2]. Однако, несмотря на значительный прогресс в исследовании микроРНК, до сих пор остается множество неизведанных аспектов их взаимодействия с микробиомом, что требует дальнейшего изучения. Целью данного исследования является углубленное изучение роли микроРНК в регуляции кишечного микробиома человека и идентификация бактериальных генов-мишеней, которые могут быть задействованы в этом процессе, с использованием биоинформационных баз данных и онлайн-ресурсов. Это позволит не только расширить понимание молекулярных механизмов взаимодействия микроРНК и микробиома, но и способствовать разработке новых стратегий для коррекции микробиоты и улучшения здоровья.

Методы и алгоритмы: Для поиска интересующих нас микроРНК были использованы базы данных: miRbase (https://www.mirbase.org/), miRDB (https://mirdb.org/), TargetScanHuman (https://www.targetscan.org/vert_80/). Для предсказания генов-мишеней были использованы: IntaRNA (http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp) и TargetRNA3 (https://cs.wellesley.edu/~btjaden/TargetRNA3/).

Результаты: Из литературных источников были отобраны 10 предполагаемых микроРНК с последовательностями, которые влияют на кишечный микробиом человека (miR-515-5p, miR-1226-5p, miR-194-5p, miR-876-5p, miR-148-3p, miR-21, miR-101, miR-325, miR-623, miR-1253) [3]. С помощью базы данных miRbase.org были получены последовательности интересующих нас 10 пре-микроРНК, после чего была сделана визуализация шпилек пре-микроРНК с использованием веб-сервиса ViennaRNA(http://www.viennarna.at/). Для всех микроРНК прокариот были предсказаны гены-мишени с помощью инструмента IntaRNA. В итоге были получены большое количество предполагаемых генов мишеней (приблизительно по 100 генов для каждой микроРНК). Вследствие обилия генов мы решили выделить общие мишени среди исследуемых микроРНК для рассмотрения механизмов воздействия на представителей кишечного сообщества человека.

Несмотря на различную природу исследуемых микроРНК, можно предположить схожесть их генов-мишеней, что указывает на схожий спектр влияния. При использовании TargetRNA3 были предсказаны гены-мишени только для 5 микроРНК (miR-1226-5p, miR-194-5p, miR-148-3p, miR-101, miR-1253). Всего отобрано 4 гена – *mukB*, *cysI*, *fadK* и C7Y58_RS04465. Больше всех был предказан ген *mukB*, для miR-1226-5p, miR-194-5p, miR-148-3p и имел большую энергию связи с данными микроРНК: miR-1226-5p – –8.34 kcal/mol, miR-194-5p – –5.33 kcal/mol, miR-148-3p – –8.15 kcal/mol. Продукт данного гена играет центральную роль в конденсации хромосом, сегрегации и развитии клеточного цикла. Функционирует как гомодимер, который необходим для разделения хромосом. Участвует в отрицательной суперспирализации ДНК *in vivo* и тем самым организует и уплотняет хромосомы. Может достигать или облегчать сегрегацию хромосом за счет конденсации ДНК с обеих сторон центрально расположенной реплисомы во время клеточного деления. Стимулирует как релаксацию ДНК, так и, в меньшей степени, активность декатенации топоизомеразы IV [4].

Выводы: В рамках настоящего исследования был проведен анализ потенциальных генов-мишеней микроРНК в контексте взаимодействия человека и кишечного микробиома с использованием биоинформатических инструментов IntaRNA и TargetRNA3. Полученные данные указывают на значительные различия в предсказаниях генов-мишеней, что подчеркивает сложность молекулярных взаимодействий микроРНК с ними, а также необходимость дальнейшего углубленного анализа и экспериментальной валидации. Стоит упомянуть роль гена mukB, имевшего большую энергию связи с микроРНК (miR-1226-5p, miR-194-5p, miR-148-3p): ген mukB, предсказанный как мишень для нескольких микроРНК, играет центральную роль в жизненно важных процессах бактериальной клетки, что подчеркивает потенциальное значение микроРНК в регуляции клеточного цикла и поддержании структуры генома микроорганизмов. Дальнейшие исследования данного гена могут позволить лучше понять процесс взаимодействия микробиоты и хозяина, а также разработать новые методы микроРНК-терапии заболеваний ЖКТ (желудочно-кишечного тракта), на данный момент очень перспективным является использование антисмысловых олигонуклеотидов в лечении различных патологий [5]. Результаты подчеркивают необходимость комплексного подхода к анализу взаимодействий микроРНК и микробиоты ЖКТ, включая использование различных биоинформатических инструментов, экспериментальную валидацию и обновление геномных баз данных. Дальнейшие исследования должны сосредоточиться на углубленном понимании этих взаимодействий для разработки новых стратегий профилактики и лечения заболеваний, связанных с микробиотой ЖКТ.

Финансирование: Исследование поддержано РНФ (грант 24-24-00563, «Разработка цифровых образовательных программ в биомедицине»).

Search for microRNA target genes in human-gut microbiome interactions

Troshina D.A.[2]*, Shoshin F.V.[3], Stepanova A.A.[1], Orlov Y.L.[1, 4]

[1] *First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

[2] *Lomonosov Moscow State University, Moscow, Russia*

[3] *YSMU of the Ministry of Health of Russia, Yaroslavl, Russia*

[4] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

** darya.troshina02@mail.ru, orlov@bionet.nsc.ru*

Key words: medical informatics; microRNA target genes; intestinal microbiome; microRNA; IntaRNA; TargetRNA3

Motivation and Aim: Modern medicine faces the need to develop new approaches to the prevention and treatment of a wide range of diseases associated with an imbalance of the intestinal microbiome. MicroRNAs, as key regulators of gene expression [1], represent promising molecular tools to modulate the microbiota [2]. However, despite the significant progress in the study of microRNAs, there are still many unexplored aspects of their interaction with the microbiome that require further investigation. The aim of this study is to investigate the role of microRNAs in the regulation of the human intestinal microbiome and to identify bacterial target genes that may be involved in this process using bioinformatics databases and online resources. This will not only increase the understanding of the molecular mechanisms of microRNA-microbiome interactions, but also contribute to the development of new strategies to correct the microbiota and improve health.

Methods and Algorithms: The following databases were used to search for microRNAs of interest: miRbase (https://www.mirbase.org/), miRDB (https://mirdb.org/), and TargetScanHuman (https://www.targetscan.org/vert_80/). IntaRNA (http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp) and TargetRNA3 (https://cs.wellesley.edu/~btjaden/TargetRNA3/) were used to predict target genes.

Results: We selected 10 putative microRNAs with sequences that affect the human intestinal microbiome (miR-515-5p, miR-1226-5p, miR-194-5p, miR-876-5p, miR-148-3p, miR-21, miR-101, miR-325, miR-623, miR-1253) from the literature [3]. The sequences of the 10 pre-miRNAs of interest were obtained using the miRbase.org database, followed by visualisation of the pre-miRNA hairpins using the ViennaRNA web service (http://www.viennarna.at/). For all prokaryotic microRNAs, target genes were predicted using the IntaRNA tool. As a result, a large number of putative target genes were obtained (approximately 100 genes for each microRNA). Due to the abundance of genes, we decided to identify common targets among the studied microRNAs in order to consider the mechanisms of impact on the human gut community. Despite the different nature of the studied microRNAs, we can assume the similarity of their target genes, indicating a similar spectrum of influence. Using TargetRNA3, target genes were predicted for only 5 microRNAs (miR-1226-5p, miR-194-5p, miR-148-3p, miR-101, miR-1253). A total of 4 genes, mukB, cysI, fadK and C7Y58_RS04465, were selected. The mukB gene was most predicted for miR-1226-5p, miR-194-5p, miR-148-3p and had a high binding energy to these miRNAs: miR-1226-

5p – –8.34 kcal/mol, miR-194-5p – –5.33 kcal/mol, miR-148-3p – –8.15 kcal/mol. The product of this gene plays a central role in chromosome condensation, segregation and cell cycle progression. Functions as a homodimer that is required for chromosome segregation. Participates in the negative superhelicalisation of DNA *in vivo* and thereby organises and compacts chromosomes. May achieve or facilitate chromosome segregation by condensing DNA on both sides of a centrally located replisome during cell division. Stimulates both DNA relaxation and, to a lesser extent, topoisomerase IV decatenation activity [4].

Conclusion: This study analysed potential microRNA target genes in the context of human-gut microbiome interactions using the bioinformatic tools IntaRNA and TargetRNA3. The data obtained indicate significant differences in the prediction of target genes, which emphasises the complexity of molecular interactions of microRNAs with them, as well as the need for further in-depth analysis and experimental validation. It is worth mentioning the role of the mukB gene, which had a high binding energy with microRNAs (miR-1226-5p, miR-194-5p, miR-148-3p): the mukB gene, predicted as a target for several miRNAs, plays a central role in vital processes of the bacterial cell, highlighting the potential importance of microRNAs in cell cycle regulation and maintenance of microbial genome structure. Further studies of this gene may allow us to better understand the process of interaction between microbiota and host, as well as to develop new methods of microRNA therapy of GI (gastrointestinal tract) diseases. At the moment, the use of antisense oligonucleotides in the treatment of various pathologies is very promising [5]. The results emphasise the need for a comprehensive approach to the analysis of interactions between miRNAs and the GI microbiota, including the use of various bioinformatic tools, experimental validation and updating of genomic databases. Further research should focus on a deeper understanding of these interactions to develop new strategies for the prevention and treatment of diseases related to the GI microbiota.

Funding: This work was supported by Russian Science Foundation (grant project 24-24-00563, «Development of digital education programs in biomedicine»).

Список литературы/References

1. Saini H.K., Enright A.J., Griffiths-Jones S. Annotation of mammalian primary microRNAs. *BMC Genomics*. 2008;9:564
2. Yu Y. et al. MicroRNAs: The novel mediators for nutrient-modulating biological functions. *Trends Food Sci Technol*. 2021;114:167-175
3. Bi K., Zhang X., Chen W., Diao H. MicroRNAs Regulate Intestinal Immunity and Gut Microbiota for Gastrointestinal Health: A Comprehensive Review. *Genes*. 2020;9:1075
4. Sawitzke J., Austin S. An analysis of the factory model for chromosome replication and segregation in bacteria. *Mol Microbiol*. 2001;40:786-794
5. Gebert L.F. et al. Miravirsen (SPC3649) can inhibit the biogenesis of miR-122. *Nucleic Acids Res*. 2014;42(1):609-621

doi 10.18699/bgrs2024-9.1-15

Assigning transcriptomic subtypes to CML/CLL samples using nanopore RNA-sequencing and self-organizing maps

Arakelyan A.[1, 2, 3]*, Hakobyan Y.[4], Sirunyan T.[1, 3], Khachatryan G.[1, 3], Ghukasyan L.[1], Minasyan A.[1], Hakobyan S.[2,5], Martirosyan G.[1], Chavushyan A.[1], Binder H.[5,6].

[1] *Bioinformatics Group, Institute of Molecular Biology NAS RA, Yerevan, Armenia*

[2] *Laboratory of Human Genomics, Institute of Molecular Biology NAS RA, Yerevan, Armenia*

[3] *Institute of Biomedicine and Pharmacy, Russian-Armenian University, Yerevan, Armenia*

[4] *Hematology Center after Prof. Yeolyan, Hematology and Transfusion Medicine department of NIH Armenia, Yerevan, Armenia*

[5] *Armenian Bioinformatics Institute, Yerevan, Armenia*

[6] *Interdisciplinary Center for Bioinformatics, Leipzig University, Leipzig, Germany*

* *arakelyan@sci.am*

Key words: nanopore RNA-sequencing; CML; CLL; self-organizing maps; support vector machines; transcriptomic subtypes; RNA sequencing data analysis

Motivation and Aim: Next-generation sequencing technologies have revolutionized the fields of cancer screening, diagnostics, and precision oncology for blood cancers [1]. While Illumina platforms offer robust performance, they require significant initial infrastructure investments, and a high volume of samples for pooling, and may exhibit extended turnaround times when sample numbers are low. Conversely, third-generation (long-read) sequencing from Oxford Nanopore Technologies (ONT) can substantially reduce sequencing costs, making it a valuable tool for cost-effective and accurate cancer testing [2]. However, nanopore sequencing is challenged by lower accuracy and sample throughput compared to Illumina platforms. To enhance its utility, additional computational strategies are necessary. In our study, we integrated publicly available Illumina RNA sequencing data with machine learning methods to classify samples of chronic myeloid leukemia (CML) and chronic lymphocytic leukemia (CLL) into transcriptomic subtypes using nanopore RNA sequencing data.

Methods and Algorithms: Publicly available Illumina RNA sequencing data were obtained from the Gene Expression Omnibus (for CML, accession: GSE100026) and cBioPortal (for CLL, https://www.cbioportal.org/study/summary?id=cll_broad_2015). Gene counts in RPKM format were converted to TPM values. We then utilized the oposSOM R package for portraying high-dimensional data using self-organizing maps (SOM) [3]. This "SOM portrayal" allowed us to create transcriptome maps for CML and CLL, and identify co-expressed gene modules, and transcriptomic subtypes.

Twenty-five patients (six with CML and nineteen with CLL) were recruited from the Hematology Center After Prof. R. Yeolyan. Total RNA was isolated from blood samples using the Quick DNA/RNA Miniprep Plus Kit (Zymo Research, US). Sequencing libraries were prepared using the PCR-cDNA sequencing-barcoding kit (SQK-PCB109, Oxford Nanopore Technologies, UK) according to the manufacturer's instructions, loaded onto an R9.4.1 flow cell, and sequenced using a MinION mk1b instrument. The sequencing run lasted between 24 and 48 hours. Raw sequencing data were basecalled, demultiplexed, and converted to Fastq files using the Guppy tool. Reads were then

aligned to the human reference genome (hg38) using the Minimap2 software. Read counts were estimated using the featureCounts function from the Subread R package, based on GENCODE v44 annotations. Raw reads were transformed to TPM. Subsequently, TPM data from the patients were projected into the SOM space generated with Illumina RNA-seq data using a supervised SOM portrayal (supSOM) approach [4], which integrates a support vector machine regression model with the original SOM algorithm to predict the SOM portrait of a new sample.

The study received approval from the Ethics Committee of the Institute of Molecular Biology NAS RA.

Results: The SOM portrayal of CML enabled the identification of gene expression modules associated with the chronic (CP) and blast crisis (BP) phases. Samples from the CP phase exhibited upregulation of gene sets linked to oxidative phosphorylation, lipid metabolism, and proliferation, along with downregulated immune gene signatures. In the BP phase, there was noticeable upregulation of the RB1 pathway, angiogenesis, and inflammation gene signatures. The projection of ONT RNA sequencing data revealed significant differences from normal peripheral blood mononuclear cells (PBMCs) in the SOM space, with upregulation in gene signatures common to both CP and BP samples. These findings suggest that ONT CML samples are transitioning from the chronic to the blast crisis phase.

In CLL, the SOM space divided samples into eight transcriptomic subtypes (A *, A G *, A K J F * – immune phenotype, G P * – inflammatory, K *, M *, N * – non-immune, P I *, P L * – angioplastic), each characterized by distinct marker gene signatures. The projection of ONT CLL samples onto this SOM space indicated that the samples belonged to the M * and N * transcriptomic subtypes, which are characterized by low expression of ZAP70 and CD38 and associated with a favorable prognosis.

Conclusion: Our findings demonstrate that shallow ONT sequencing, when integrated with publicly available data and machine learning algorithms, can facilitate molecular subtyping CML and CLL. The ability to classify these cancers into subtypes can be beneficial for tailoring treatment strategies and improving patient management.

Funding: This study was funded by a research grant from the Committee of Higher Education and Science of the Ministry of Education and Science of the Republic of Armenia (21AG-1F021, PI: AA).

References

1. Schwartzberg L., Kim E.S., Liu D., Schrag D. Precision Oncology: Who, How, What, When, and When Not? *Am Soc Clin Oncol Educ Book*. 2017;37:160-169. doi 10.1200/EDBK_174176
2. Wang Y., Zhao Y., Bollas A., Wang Y., Au K.F. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol*. 2021;39(11):1348-1365. doi 10.1038/s41587-021-01108-x
3. Löffler-Wirth H., Kalcher M., Binder H. oposSOM: R-package for high-dimensional portraying of genome-wide expression landscapes on bioconductor. *Bioinformatics*. 2015;31(19):3225-3227. doi 10.1093/bioinformatics/btv342
4. Nikoghosyan M., Loeffler-Wirth H., Davidavyan S., Binder H., Arakelyan A. Projection of High-Dimensional Genome-Wide Expression on SOM Transcriptome Landscapes. *BioMedInformatics*. 2022;2:62-76. doi 10.3390/biomedinformatics2010004

doi 10.18699/bgrs2024-9.1-16

Single-cell transcriptomic analysis reveals a unique cluster and differences in gene expression in somatotroph adenomas

Asaad W.[1]*, Deviatiiarov R.[1, 2, 3, 4], Shcherbakova A.[1], Popov S.[1], Utkina M.[1]**

[1] *Department of General, Molecular and Population genetics, Endocrinology Research Centre, Moscow, Russia*

[2] *Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[3] *Graduate School of Medicine, Juntendo University, Tokyo, Japan*

[4] *Life Improvement by Future Technologies (LIFT) Center, Moscow, Russia*

** walaaakasaad94@gmail.com; ** mv.utkina@yandex.ru*

Key words: Silent somatotroph adenoma; functioning somatotroph adenoma; pituitary adenomas; scRNA-seq; PitNETs

Motivation and Aim: Pituitary adenomas (PAs) are benign tumors of the pituitary gland, accounting for approximately 15 % of all primary brain tumors [1]. They can be either functioning (secreting) or non-functioning (silent). Functioning somatotroph adenoma (FSA) is a type of PAs that primarily secretes growth hormone (GH). The silent form of somatotroph adenomas (SSAs) does not secrete hormones and is more invasive and frequent than its secreting counterparts [2]. In our study, we aim to describe SSAs using single-cell RNA sequencing (scRNA-seq) for the first time. We will compare them with scRNA-seq data of FSAs, also included in our study, to understand the differences in the molecular mechanisms regulating gene expression in each tumor type.

Methods and Algorithms: We obtained pituitaries from four patients (three are FSAs and one SSA), and subjected these tissues to scRNA-seq with a total of 15176 and 4294 cells respectively. The raw sequenced reads were processed with 10X Cell Ranger (v6.1.1), and Default Cell Ranger quality check measurements were used for further comparison methods through the Wilcoxon test. The expression matrices for the filtered cells were submitted to Seurat (v4.9.9 and v5.0.0) for basic analysis, including scaling and normalization. Cell cycle phase predictions were based on reference gene expression processed with Seurat (v5.0.0). Statistical significance was assessed by a two-tailed t-test and Wilcoxon rank-sum test: * ($0.01 < p < 0.05$), ns – not significant – $p > 0.05$.

Results: Based on known tissue-specific marker genes, we identified six main cell clusters in both FSA and SSA samples (Fig. 1a). These cell clusters include the PIT1 lineage (somatotrophs, thyrotrophs, and lactotrophs), the TPIT lineage (corticotrophs), the SF1 lineage (gonadotrophs), stroma cells (endothelial and perivascular cells), immune cells (macrophages and T-cells), and stem cells. The hormone-secreting cells were identified within each lineage based on the enriched gene expression levels of well-known markers. We compared the gene expression profile between FSAs and SSAs. We found that FSAs differentially expressed *SSTR1*, *SSTR5*, and *NPY* genes that code for hormone receptors and secretion regulatory proteins, which correlates with the secreting properties of these tumors rather than the non-secreting SSAs (Fig. 1d).

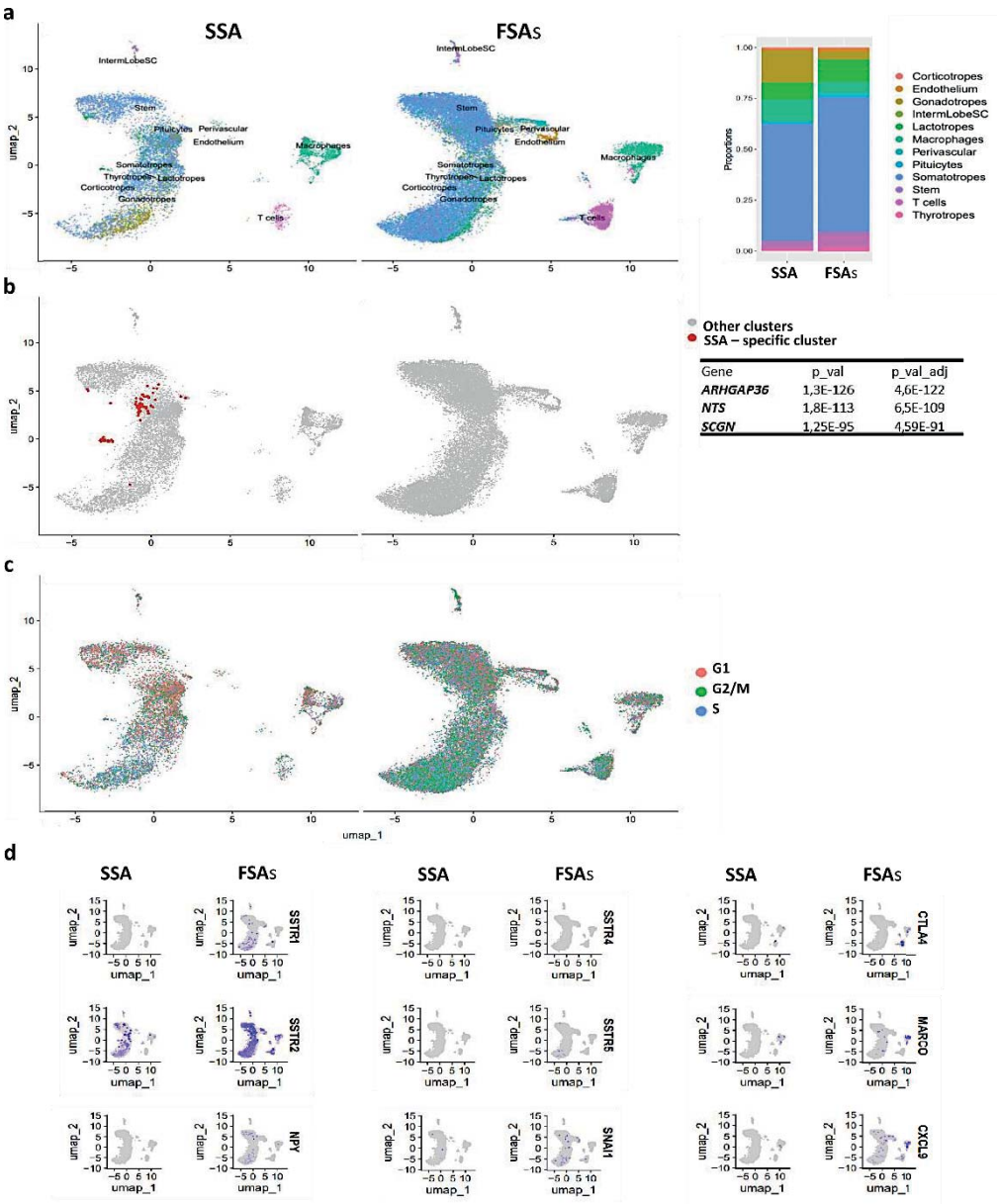


Fig. 1. scRNA-seq data analysis for FSA and SSA samples. *a*: The major cell type compositions found in FSA and SSA samples are represented as UMAP on the left, and a bar plot on the right. *b*: The unique SSA – specific cluster is shown in red, existing only in SSA and not in FSA samples. Examples of differentially expressed (DE) genes associated with the SSA – specific cluster are shown in the table. *c*: Cell cycle estimation for FSA and SSA datasets. *d*: Key specific gene expression according to UMAP for SSA and FSAs cells

FSAs also differentially expressed *SNAI1* – an epithelial to mesenchymal transition (EMT) inducer gene, and immune cell-specific genes (*CTLA4, CXCL9*, and *MARCO*) which indicate a difference in the tumor microenvironment and immune profile between FSAs and SSAs (Fig. 1d). Additionally, we detected a novel cluster (SSA-specific cluster) (Fig. 1b), which is found in SSAs but not in FSAs. This cluster shows high activation of cell pathways involved in cell growth, cell differentiation, cell migration, apoptosis, and tumor development like PI3K/AKT/mTOR, JAK/STAT, and TGFB signaling pathways [3–5]. This cluster also highly expresses the *NTS* gene, which is correlated with poor tumor prognosis [6], and genes of ribosomal proteins which might to be involved in cancer initiation [7]. These findings suggest that this cluster plays a significant role in the pathogenesis of SSAs. Furthermore, our scRNA defined a difference in the overall cell cycle status of FSAs and SSAs, where FSAs mostly locate in the G2M phase while SSAs are in the G1 and S phases (Fig. 1c). This reflects a difference in the molecular mechanisms regulating tumor cell growth and proliferation between these two tumor types.

Conclusion: In conclusion, our study of secreting and silent somatotroph adenomas revealed distinct cell clusters in both adenoma types. Furthermore, differences in gene expression related to immune response and tumor microenvironment were observed between functioning and silent adenomas. A unique cluster found only in silent adenomas exhibited high activation of pathways associated with cell growth, differentiation, migration, and tumor development, suggesting a pro-tumor phenotype. Additionally, our data indicate variations in molecular mechanisms regulating cell cycle status in SSAs and FSAs. Overall, these findings provide valuable insights into the distinct characteristics and potential prognostic factors of functioning and silent pituitary adenomas.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-15-2022-310 from 20 April 2022).

References

1. Ostrom Q.T. et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014-2018. *Neuro Oncol*. 2021;23(12 Suppl. 2):iii1-iii105
2. Naritaka H. et al. Morphological characterization and subtyping of silent somatotroph adenomas. *Pituitary*. 1999;1(3-4):233-241
3. Hu Q. et al. JAK/STAT pathway: Extracellular signals, diseases, immunity, and therapeutic regimens. *Front Bioeng Biotechnol*. 2023;11:1110765
4. Zhang Y., Alexander P.B., Wang X.F. TGF-β Family Signaling in the Control of Cell Proliferation and Survival. *Cold Spring Harb Perspect Biol*. 2017;9(4):a022145
5. Peng Y. et al. PI3K/Akt/mTOR Pathway and Its Role in Cancer Therapeutics: Are We Making Headway? *Front Oncol*. 2022;12:819128
6. Christou N. et al. Neurotensin pathway in digestive cancers and clinical applications: an overview. *Cell Death Dis*. 2020;11(12):1027
7. Goudarzi K.M., Lindström M.S. Role of ribosomal protein mutations in tumor development (Review). *Int J Oncol*. 2016;48(4):1313-1324

doi 10.18699/bgrs2024-9.1-17

Differential expression of various LncRNAs and microRNAs in abdominally obese and non-obese individuals

Bairqdar A.[1, 3]*, Ivanoshchuk D.[1, 2], Shirokova N.[1], Tuzovskaya O.[2], Kashtanova E.[2], Polonskaya Y.[2], Shakhtshneider E.[1, 2]

[1] *Federal Research Center, Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia*

[2] *Institute of Internal and Preventive Medicine, Branch of Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia*

[3] *Department of Genetics, Novosibirsk State University, Novosibirsk, Russia*

* *bairqdar@bionet.nsc.ru*

Key words: Obesity; abdominal obesity, miRNA; lncRNA; visceral adipose tissue

Motivation and Aim: Obesity is a complex, multifactorial disease characterized by excessive fat accumulation that may impair health [1]. Despite significant advances in our understanding of obesity, many aspects of its underlying molecular mechanisms remain elusive. Long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) are increasingly explored as critical regulators or biomarkers in various diseases [2], with their roles in adipogenesis, lipid metabolism, and insulin response currently being studied [3]. However, studies linking lncRNAs and miRNAs to obesity are relatively scarce and sometimes yield inconsistent results [4]. Our research investigates the differential expression of a wide range of lncRNAs (ASMER1, SNHG9, P5549, p19461, GAS5 and MALAT1) and miRNAs (MIR26A1, MIR222, MIR221 and MIR155) in the visceral adipose tissues of individuals with and without abdominal obesity.

Methods and Algorithms: Population and Sampling: The study included a case group with abdominal obesity and a control group without obesity, in total 100 participants. Abdominal obesity was defined accordingly Body Mass Index (BMI) ≥ 25.0 kg/m² for whites, and for Asians BMI ≥ 23.0 kg/m² with a waist circumference (WC) ≥ 80 cm for women and WC ≥ 94 cm for men [5]. White visceral adipose tissue samples were collected from individuals during elective surgery.

RNA Extraction: Tissue samples were preserved immediately in RNAlater (Thermo Fisher Scientific, USA) and stored at –20 °C until RNA extraction. RNA was extracted and cDNA was synthesized according to manufacturers protocols (Biolabmix, Russia). Quantitative PCR (qPCR) was performed using SYBR dye (Biolabmix, Russia) to determine the relative expression levels of targeted microRNAs and lncRNAs using the LightCycler96 system (Roche, Switzerland). The ΔΔCq method was utilized for expression level calculations.

Statistical Analysis: The normality of data distribution and homogeneity of variances were assessed using Levene's test. Correlations between RNA levels and various clinical indicators were determined using Student's t-tests and Mann–Whitney U tests. The Bonferroni correction was applied to adjust for multiple comparisons.

Results: Expression of ASMER1 and MALAT1 as well as miR222 and miR221 was undetectable in the homogenate of visceral fat tissue. RNA levels of P5549 and P19461 were at the borderline of detection levels and showed no reliable differences between the control and case groups. Expression levels of SNHG9, GAS5, and miR155 showed

no significant difference between participants with and without obesity, while the expression of miR26A1 was significantly lower ($P<0.01$) in the case group.

Conclusion: Our study is the first to show that miR26A is significantly downregulated in the visceral adipose tissue of obese individuals. This finding is consistent with previously described functions of miR26A, such as suppressing adipocyte progenitor differentiation and reducing visceral fat mass and blood lipid levels in mice [6].

In contrast, GAS5 levels remained nearly unchanged between groups and showed no correlation with miR26A levels. This result is inconsistent with multiple studies that suggest a regulatory 'sponge' interaction between GAS5 and miR26A [7]. This discrepancy suggests that an alternative mechanism may be responsible for the downregulation of miR26A in human visceral adipose tissue, necessitating further investigation.

Funding: The molecular genetic testing was carried out within the framework of the main topic of state assignment No. FWNR-2022-0003.

References

1. Inoue Y., Qin B., Poti J., Sokol R., Gordon-Larsen P. Epidemiology of Obesity in Adults: Latest Trends. *Curr Obes Rep*. 2018;7(4):276-288. doi 10.1007/s13679-018-0317-8
2. Nemeth K., Bayraktar R., Ferracin M. et al. Non-coding RNAs in disease: from mechanisms to therapeutics. *Nat Rev Genet.* 2024;25:211-232. doi 10.1038/s41576-023-00662-1
3. Sufianov A., Beilerli A., Kudriashov V. et al. The role of long non-coding RNAs in the development of adipose cells. *Noncoding RNA Res*. 2023;8(2):255-262. doi 10.1016/j.ncrna.2023.02.009
4. Corral A., Alcala M., Carmen Duran-Ruiz M. et al. Role of long non-coding RNAs in adipose tissue metabolism and associated pathologies. *Biochem Pharmacol*. 2022;206:115305. doi 10.1016/j.bcp.2022.115305
5. Dedov I.I., Shestakova M.V., Melnichenko G.A. et al. Interdisciplinary Clinical Practice Guidelines "Management of Obesity and Its Comorbidities". *Obes Metab*. 2021;18:5-99
6. Zeng H., Sun W., Ren X. et al. AP2-microRNA-26a overexpression reduces visceral fat mass and blood lipids. *Mol Cell Endocrinol*. 2021;528:111217. doi 10.1016/j.mce.2021.111217
7. Tan L., Xie Y., Yuan Y., Hu K. LncRNA GAS5 as miR-26a-5p Sponge Regulates the PTEN/PI3K/Akt Axis and Affects Extracellular Matrix Synthesis in Degenerative Nucleus Pulposus Cells in vitro. *Front Neurol*. 2021;12:653341. doi 10.3389/fneur.2021.653341

doi 10.18699/bgrs2024-9.1-18

Distribution of epidermal proliferating cells in the model of human skin xenograft

Cherkashina O.[1]*, Tsitrina A.[2], Kosykh A.[3], Vorotelyak E.[1], Kalabusheva E.[1]

[1] *Koltzov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, Russia*

[2] *Ilse Katz Institute of Nanoscale Science, Beer Sheva, Israel*

[3] *Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Pirogov Russian National Research Medical University, Moscow, Russia*

* *olgalcher@gmail.com*

Key words: skin regeneration; epidermal proliferation; skin stem cells; skin xenograft

Motivation and Aim: The study of proliferation is of great importance for the comprehension of the processes underlying skin regeneration in the context of wound healing and pathological conditions, as well as for the understanding of normal skin functioning. A plethora of models has been developed to predict the pattern of cell divisions in the skin. However, despite this, the question of the precise pattern of epidermal proliferation remains controversial. Some models propose the existence of a single-progenitor whose asymmetrical divisions result in the generation of other progenitor cells or differentiating cells. In the initial studies, epidermal proliferation units were described, comprising approximately 10 basal cells with a clonogenic cell in the center and suprabasal cells located above them [1, 2]. Later, it has been proposed that cells in the epidermis are evenly distributed, with a more random division pattern. In this scenario, cell localization at epidermal rete ridges or inter-ridges may play a pivotal role. It has been shown that human rete ridges are transcriptionally similar to mouse tail scales, which are characterized by a higher cell division rate [3]. However, epidermal cell proliferation in human skin is less well-studied than in mice due to the difficulty of utilizing vital labels in humans, which is complicated by bioethical considerations.

The objective of this study was to analyze the dynamics of cell proliferation using a xenotransplantation model, which allows for the modeling of the regeneration process.

Methods and Algorithms: To investigate the process of skin regeneration, we developed a model of human skin xenograft transplantation into mice with immunodeficiency. Biopsies were obtained at 40, 75, and 110 days post-transplantation, fixed in optimal cutting temperature (O.C.T.) compound, and cryosectioned. To study cell proliferation, a bromodeoxyuridine (BrdU) label was injected intraperitoneally seven days prior to sacrifice. The cryosections were then immunohistochemically stained with Ki67 and BrdU, imaged at high resolution, and analyzed by the QuPath-0.5.0 program.

A QuPath algorithm was developed to study the distribution of proliferating cells. The StarDist extension was used for nuclei segmentation based on the DAPI channel, with subsequent classification of detected nuclei based on Ki67/BrdU labeling. Nuclei were classified as negative, Ki67+BrdU-, Ki67-BrdU+, or Ki67+BrdU+. Additionally, BrdU+ cells were classified based on BrdU fluorescence intensity. All cells were divided by their localization in the basal or suprabasal epidermal layer. To analyze the cell distribution pattern, the epidermis was segmented into annotations containing two to four basal cells and located above them suprabasal cells. K-means cluster analysis was

performed to classify these annotations. The output data was analyzed with Microsoft Excel 2016 and Origin 2022.

Results: Following xenografting, the structure of human skin is gradually restored, and by day 110, epidermal morphology and marker distribution is close to normal skin, thus rendering it suitable for use as a healthy skin model. To analyze cell proliferation, Ki67 staining is a common marker of cells in the active phase of the cell cycle. The introduction of BrdU prior to the sacrifice allows for labeling of cells that have divided following the BrdU injection. High intensity of the BrdU staining is indicative of cells that have divided only once. During subsequent divisions, the intensity of the BrdU label decreases. Based on these data, cells may be classified by their division rate and the data may be used to analyze cell proliferation patterns. In order to conduct a cluster analysis of dividing epithelial cells, image annotations were created that included 2-4 basal and overlying suprabasal cells. The number of cells included in each annotation corresponds to an approximate size of an epithelial proliferation unit. The analysis of these annotations and their clustering will provide valuable insights into the distribution of proliferating cells. The Ki67+BrdU+ cells were observed in both rete-ridges and inter-ridges, with a higher density in the rete ridges. Our observations are consistent with previous studies that have identified the rete ridges as a site of active epidermal proliferation. The distribution of Ki67-BrdU+ cells, which had undergone one or more divisions and entered the G0 phase of the cell cycle, was observed to differ from that of other cell groups in that these cells were distributed at a greater distance from the basal membrane. At the early stage of regeneration, the xenograft displays a thickened hyperproliferating epidermis. Following xenografting, the epidermal layers exhibited a thickening with a delayed differentiation of basal cells. The number of proliferating Ki67+ and BrdU+ cells increased significantly, with Ki67+BrdU+ cells being distributed over a larger distance from the basal membrane. As the skin xenograft gradually restores its structure, it may serve as a model for studying skin regeneration, with late stages corresponding to normal skin morphology.

Conclusion: An algorithm has been developed for the analysis of epidermal proliferation patterns. As previously demonstrated, human rete ridges exhibit a closer transcriptional relationship with mouse scales, which are distinguished by their elevated proliferation rates [3]. The use of human skin xenografts permitted the investigation of cell division rates in human epidermis through the use of BrdU tracking. The presence of Ki67+BrdU+ cells was observed in both rete ridges and inter-ridges, with a higher density observed at the former. This finding supports the functional similarity of these human and mouse epidermal domains.

Funding: The study is supported by the Russian Science Foundation (project No. 21-74-30015, https://rscf.ru/en/project/21-74-30015/).

References

1. Roy E., Neufeld Z. et al. Bimodal Behaviour of Interfollicular Epidermal Progenitors Regulated by Hair Follicle Position and Cycling. *EMBO J.* 2016;35:2658-2670. doi 10.15252/embj.201693806
2. Sada A., Jacob F., Leung E. et al. Defining the Cellular Lineage Hierarchy in the Interfollicular Epidermis of Adult Skin. *Nat Cell Biol.* 2016;18:619-631. doi 10.1038/ncb3359
3. Ghuwalewala S., Lee S.A., Jiang K. et al. Epidermal Basal Domains Organization Highlights Skin Robustness to Environmental Exposure. *bioRxiv*. 2022. doi 10.1101/2022.02.23.481662
4. Schmidt U., Weigert M., Broaddus C., Myers G. Cell Detection with Star-Convex Polygons. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Springer, 2018;265-273. doi 10.1007/978-3-030-00934-2_30
5. Bankhead P., Loughrey M.B., Fernández J.A. et al. QuPath: Open Source Software for Digital Pathology Image Analysis. *Sci Rep.* 2017;7:16878. doi 10.1038/s41598-017-17204-5

doi 10.18699/bgrs2024-9.1-19

Knockout or deletion of the *UBE2A* gene leads to disruption of the Rho/ROCK signaling pathway and impaired migration of neural cells differentiated from iPSCs

Fedorenko A.[1]*, Sekretova E.[1], Khomyakova E.[1], Surdina A.[1], Smirnov I.[2], Lebedev I.[3], Lagarkova M.[1], Bogomazova A.[1, 3]

[1] *Lopukhin Federal Research Clinical Center of Physical and Chemical Medicine, Department of Cell Biology, Moscow, Russia*

[2] *Lopukhin Federal Research Clinical Center of Physical and Chemical Medicine, Department of Structure and Function of Biopolymers, Moscow, Russia*

[3] *Tomsk National Research Medical Center, the Research Institute of Medical Genetics, Department of Ontogenetics, Tomsk, Russia*

* *afedorenko00@gmail.com*

Key words: UBE2A gene; X-linked Intellectual Disability type Nascimento; Rho/ROCK signalling pathway; Transcriptomics

Motivation and Aim: The X-linked intellectual disability Nascimento syndrome was first described by Rafaella Nascimento in 2006 [1]. It is characterized by delayed intellectual and cognitive development, speech impairment, and morphological abnormalities. The syndrome is associated with loss-of-function mutations in the *UBE2A* gene encoding ubiquitin-conjugating enzyme E2A, which participates in post-replicative DNA repair, regulation of gene transcription, and mitophagy. Additionally, duplication of the *UBE2A* gene has detrimental effects on mental development [2]. However, the precise role of the *UBE2A* gene and its dosage in neurogenesis remains to be elucidated.

Methods: To identify the cellular processes affected by the *UBE2A* gene dosage abnormalities in neural cells, we provided the transcriptomic analyses of neuronal precursor cells (NPCs) derived from human induced pluripotent stem cells (iPSCs). Our cell lines included iPSCs with knockout and overexpression of the *UBE2A* gene, and iPSCs of a patient with Nascimento syndrome caused by the deletion of the *UBE2A* gene [3]. We performed differential expression analysis and enrichment pathways analysis in the statistical environment R using the DESeq2 and clusterProfiler packages, respectively. The clusters of differentially expressed genes (DEGs) with shared expression profiles were identified using the 'DEGreport' package. We validated our transcriptomic results through qRT-PCR and proteomic analysis.

Results: We revealed more than 1500 up-regulated and 2000 down-regulated DEGs across the samples. These DEGs were clustered into 8 groups. The functional analysis of DEGs in NPCs derived from the patient's iPSCs revealed the abnormal expression of genes associated with cytoskeleton regulation, axon and dendrite development, and synapse regulation. Similar gene expression changes were observed in NPCs with the *UBE2A* knockout and, surprisingly, with the *UBE2A* gene overexpression. Among the common down-regulated DEGs, we identified 29 genes coding for plexins, ephrins, semaphorins, p21-activated kinases, and proteins of the LIM family. These genes are known as up-stream activators or down-stream effectors of the Rho/ROCK signaling pathway, which plays a critical role in neurodevelopment, including cell migration. To

analyze the NPC migration, we assessed the area of NPC migration from neurospheres using time-lapse microscopy. NPCs with the *UBE2A* knockout and overexpression exhibited a decreased migration area compared to normal NPCs. We hypothesize that impaired neuronal migration may contribute to the pathogenesis of Nascimento syndrome.

Conclusion: The Rho/ROCK signaling pathway demonstrated sensitivity to the *UBE2A* gene dosage abnormalities in NPCs differentiated from iPSCs. Specifically, knockout, deletion, or overexpression of the *UBE2A* gene led to a similar decrease in expression of Rho/ROCK signaling pathway genes in NPCs, resulting in impaired NPC migration.

Funding: The study is supported by RSF grant # 21-65-00017.

References

1. Nascimento R.M.P. et al. UBE2A, which encodes a ubiquitin-conjugating enzyme, is mutated in a novel X-linked mental retardation syndrome. *Am J Hum Genet.* 2006;79(3):549-555. doi 10.1086/507047.
2. Takenouchi T. et al. Microduplication of Xq24 and Hartsfield syndrome with holoprosencephaly, ectrodactyly, and clefting. *Am J Med Genet A.* 2012;158A(10):2537-2541. doi 10.1002/ajmg.a.35465
3. Tolmacheva E.N. et al. Delineation of clinical manifestations of the inherited Xq24 microdeletion segregating with sXCI in mothers: two novel cases with distinct phenotypes ranging from UBE2A deficiency syndrome to recurrent pregnancy loss. *Cytogenet Genome Res.* 2020;160(5):245-254. doi 10.1159/000508050

doi 10.18699/bgrs2024-9.1-20

Rare variants in *WFS1* gene in patients with young-onset diabetes

Ivanoshchuk D.[1, 2]*, Ovsyannikova A.[2], Semaev S.[1], Rymar O.[2], Shakhtshneider E.[1, 2]

[1] *Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia*

[2] *Institute of Internal and Preventive Medicine – Branch of Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia*

* *dinara@bionet.nsc.ru*

Key words: *WFS1* gene; monogenic diabetes; human genetics; next generation sequencing

Motivation and Aim: Wolfram-like syndrome or WFLS (OMIM# 614296) is a rare endocrine disorder characterized by an autosomal dominant inheritance, and it is characterized by diabetes mellitus, optic atrophy, and progressive hearing impairment. Clinically, the disease is similar to autosomal recessive Wolfram syndrome, but with a milder phenotype. Early manifestation of WFLS begins with non-autoimmune diabetes, followed later by other complications [1, 2]. The *WFS1* gene (OMIM#606201) is located on chromosome 4p16 and includes eight exons encoding the 890 amino acid protein wolframin. This protein is primarily localized in the endoplasmic reticulum [3]. It has been shown that wolframin is involved in membrane transport, regulation of calcium homeostasis in the endoplasmic reticulum. Furthermore, it seems that the protein supports the normal function of pancreatic beta cells by promoting insulin synthesis and reducing endoplasmic reticulum stress. [4]. In this study, we performed a screening for rare variants of the *WFS1* gene in patients with young-onset diabetes.

Methods and Algorithms: An analysis of genes associated with monogenic forms of diabetes was conducted in 290 individuals from Western Siberia with non-autoimmune diabetes using next-generation sequencing.

Results: We identified 2 unrelated patients with young-onset diabetes who were heterozygous for a variants of uncertain significance in the *WFS1* gene. One of this, NM_006005.3(*WFS1*):c.1124G>A (p.Arg375His) variant has been previously reported [5] in a Russian child with non-type 1 diabetes mellitus and we identified it in woman with non-autoimmune diabetes, hearing impairment, and non progressive optic atrophy. The identified variant segregated with this pathological phenotype in the proband and her father.

A novel NM_006005.3(*WFS1*): c.1336A>C (p.Ser446Arg) variant was found in a young man with autoantibody-negative diabetes and hearing loss. This variant was predicted as deleterious by *in silico* analysis (REVEL, VARITY, AlphaMissense) and was absent in gnomAD (version 3.1.2) database. According to the proband, his ancestors had diabetes mellitus for two generations, but they were not available for analysis.

Conclusion: Molecular diagnostics is a necessary step in the clinical diagnosis of monogenic forms of diabetes, especially for family screening of individuals with borderline or moderate carbohydrate metabolism disorders.

Funding: The molecular genetic testing was carried out within the framework of the main topic of state assignment No. FWNR-2022-0003.

References

1. Eiberg H. Autosomal dominant optic atrophy associated with hearing impairment and impaired glucose regulation caused by a missense mutation in the WFS1 gene. *J Med Genet*. 2005;43(5):435-440. doi 10.1136/jmg.2005.034892
2. Inoue H., Tanizawa Y., Wasson J. et al. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *Nat Genet*. 1998;20(2):143-148. doi 10.1038/2441
3. Cryns K., Sivakumaran T.A., Van den Ouweland J.M. et al. Mutational spectrum of the WFS1 gene in Wolfram syndrome, nonsyndromic hearing impairment, diabetes mellitus, and psychiatric disease. *Hum Mutat.* 2003;22:275-287. doi 10.1002/humu.10258
4. Abreu D., Asada R., Revilla J.M.P. et al. Wolfram syndrome 1 gene regulates pathways maintaining beta-cell health and survival. *Lab Invest*. 2020;100(6):849-862. doi 10.1038/s41374-020-0408-5
5. Glotov O.S., Serebryakova E.A., Turkunova M.E. et al., Whole-exome sequencing in Russian children with non-type 1 diabetes mellitus reveals a wide spectrum of genetic variants in MODY-related and unrelated genes. *Mol Med Rep*. 2019;20(6):4905-4914. doi 10.3892/mmr.2019.10751

doi 10.18699/bgrs2024-9.1-21

microRNA and response to anti-VEGF therapy of neovascular age-related macular degeneration in a Russian cohort: experience of use as a tool for personalized medicine

Kozhevnikova O.S.*, Shklyar A.A., Derbeneva A.S., Devyatkin V.A., Nikulich I.F., Fursova A.Zh.

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

** oidopova@bionet.nsc.ru*

Key words: age-related macular degeneration; microRNA; response to anti-VEGF

The leading cause of vision loss in older adults is age-related macular degeneration (AMD). AMD is a multifactorial neurodegenerative disease of the retina, the nature of which depends on many interacting factors: genetic, environmental, and epigenetic, including changes in microRNA expression patterns. MicroRNAs are small non-coding regulatory RNA molecules that modulate the expression of target genes by blocking translation through complementary binding of messenger RNAs. The freeze-thaw stability of microRNAs in plasma/serum/urine and the availability of quantitative detection methods expand the possibilities of their use as biomarkers, as well as potential mediators of physiological and pathological processes. Assessing the circulating pool of miRNAs in various biological fluids, such as blood plasma, is considered a promising approach to diagnosing AMD and assessing the effectiveness of therapy, which may contribute to early detection of the disease and monitoring of AMD progression. We conducted a study aimed at evaluating the microRNA profiles of patients divided into groups: patients with neovascular AMD who demonstrated a good response to anti-VEGF therapy, patients with a poor response, and a control group without fundus pathology. To assess miRNA expression in blood plasma, a highly sensitive profiling method on PCR panels with preadsorbed LNA primers (miRCURY, Qiagen) covering 179 miRNAs was used. The analysis showed significant variability in miRNA expression, with clustering of AMD and control groups observed. Significant changes in the level of 12 miRNAs were established. In the group of patients with a poor response in plasma, a significant increase in miR-27 and miR-23 microRNA levels was observed. It is known that these miRNAs are expressed by endothelial cells and belong to the miR-23~27~24 gene cluster responsible for the regulation of angiogenesis and the development of retinal vessels. Using bioinformatics approaches, target genes of differentially presented miRNAs and pathways regulated by them, potentially involved in the pathogenesis of neovascular AMD, have been identified. A high level of miR-27a in patients with neovascular AMD was verified by quantitative PCR with Taqman samples using an independent expanded sample. Moreover, according to optical coherent tomography data, in patients with a weak decrease less than 10 % in the CRT after 3 loading doses of the anti-VEGF drug the level of miR-27a was significantly increased compared to patients with a decrease in CRT by more than 25 % ($p = 0.039$), which confirms the potential role of this miRNA as a marker of poor response to antiangiogenic therapy.

Funding: The study is supported by RSF 21-15-00047.

doi 10.18699/bgrs2024-9.1-23

Pro-inflammatory endothelial dysfunction and systemic inflammatory response in context of modeled comorbid conditions

Shishkova D.K., Frolov A.V., Markova V.E., Markova Y.O., Kutikhin A.G.*
Laboratory for Molecular, Translational, and Digital Medicine, Research Institute for Complex Issues of Cardiovascular Diseases, Kemerovo, Russia
** antonkutikhin@gmail.com*

Key words: calcium stress; ionised calcium; calciprotein monomers; calciprotein particles; endothelial cells; monocytes; systemic inflammatory response; hepatic acute phase response

Motivation and Aim: Calciprotein particles (CPPs) are indispensable scavengers of excessive Ca^{2+} and PO_4^{3-} ions in blood, being internalised and recycled by liver and spleen macrophages, monocytes, and endothelial cells (ECs) [1, 2]. Earlier, it was shown that supraphysiological concentrations of CPPs (25 μg/mL), which have been detected in patients with end-stage renal disease, cause endothelial cell dysfunction by inducing release of pro-inflammatory cytokines (IL-6, IL-8, and MCP-1/CCL2) [3–11]. Here we evaluated whether physiological levels of CPPs (10 μg/mL), which correspond to 10 % increase in serum ionised calcium, and similar amounts of calcium incorporated into calciprotein monomers (CPM) are capable of causing endothelial activation.
Methods and Algorithms: CPM and CPPs were synthesised using supersaturation of saline solution with Ca^{2+} and PO_4^{3-} ions and albumin as a single mineral chaperone. Before the experiments, we calibrate the dose of Ca^{2+} ions, albumin-centric CPM, and albumin-centric CPPs to the physiological values observed in a human population (10 % increase in ionised calcium in a human population). Upon the addition of physiological calcium levels (10 μg/mL), delivered as ionised calcium (using $CaCl_2$ as Ca^{2+} donor), albumin-centric CPM, or albumin-centric CPPs, to human coronary artery endothelial cells (HCAEC) and human internal thoracic artery endothelial cells (HITAEC) for 24 hours, we measured gene expression by RT-qPCR and pro-inflammatory cytokines by dot blotting and ELISA. To model calcium stress in a living organism, we have intravenously administered $CaCl_2$, albumin-centric CPM, or albumin-centric CPPs (10 μg/mL) to Wistar rats.
Results: Mineral stress modeling revealed the physiological pattern of calcium distribution (free Ca^{2+} ions: 50 %; calciprotein monomers: 20 %; calciprotein particles: 30 %, unbound to bound calcium ratio of 1:1), confirming the pathophysiological relevance of CPM and CPP synthesis setting and indicating the role of CPM and CPP as primary and secondary depot of circulating Ca^{2+} ions. Addition of 10 μg/mL calcium elevated Ca^{2+} level in serum-free cell culture medium and in the rat serum by 10 % (1.1-fold) that corresponds to the difference between the highest and lowest Ca^{2+} quartiles in the human population. Hence, addition of 10 μg/mL calcium was considered as the relevant strategy to simulate the physiological increase of ionised calcium and to analyse the pathogenic effects of circulating calcium sources (Ca^{2+}, CPM, and CPP). Albumin-centric CPPs and CPM were internalised by HCAEC and HITAEC after 1-hour incubation in the pulsatile flow system. In contrast to Ca^{2+} ions and albumin-centric

CPM, albumin-centric CPPs (10 μg/mL) caused a significant increase in production of interleukin (IL)-6, IL-8, and monocyte chemoattractant protein (MCP-1/CCL2) into the milieu and elevated expression of *VCAM1*, *ICAM1*, *SELE*, *IL6*, *CXCL8*, and *CXCL1* genes by HCAEC and HITAEC after 24 hours of incubation. Further, incubation of HCAEC and HITAEC with albumin-centric CPPs induced excessive release of plasminogen activator inhibitor 1 (PAI-1/serpin E1), MCP-1/CCL2, IL-8, urokinase-type plasminogen activator receptor (uPAR), and macrophage inflammatory protein-3 alpha (MIP-3α). Co-incubation of monocytes with albumin-centric CPPs in the pulsatile flow system induce increased release of PAI-1, CXCL1 and CXCL5 chemokines, adiponectin, lipocalin-2, IL-6, IL-8, MIP-1α/1β, uPAR, MIP-3α, and matrix metalloproteinase 9 (MMP-9), whilst Ca^{2+} ions and albumin-centric CPM did not cause such consequences. However, intravenous administration of Ca^{2+} ions, albumin-centric CPM and albumin-centric CPPs (10 μg/mL calcium) to Wistar rats triggered a pronounced cytokine response documented as an elevation of 12, 17, and 14 cytokines respectively. Among these cytokines were granulocyte-macrophage colony-stimulating factor (GM-CSF), CX3CL1 (fractalkine), MCP-1/CCL2, CXCL7, CCL11, CCL17, PAI-1, MMP-2, MMP-3, hepatokines (hepassocin, fetuin A, hepatocyte growth factor (HGF), fibroblast growth factor 21 (FGF-21), growth/differentiation factor 15 (GDF-15)), receptor for advanced glycation end-products (RAGE), adiponectin, fibulin-3, galectin-1, and endostatin. Hence.

Conclusion: Among three distinct mechanisms of calcium delivery, albumin-centric CPPs were the only which provoked pro-inflammatory endothelial cell dysfunction and pro-inflammatory monocyte activation if added at physiological levels (10 μg/mL), whilst Ca^{2+} ions and CPM did not cause significant alterations. *In vivo*, calcium stress induced systemic inflammatory response and hepatic acute phase response irrespectively of the calcium source.

Funding: This research was funded by the Russian Science Foundation, grant number 22-15-00107 “Circulation of calciprotein particles in human blood: pathogenic consequences and molecular mechanisms” (Anton Kutikhin), https://rscf.ru/en/project/22-15-00107/.

References

1. Smith E.R., Hewitson T.D., Jahnen-Dechent W. Calciprotein particles: mineral behaving badly? *Curr Opin Nephrol Hypertens*. 2020;29(4):378-386. doi 10.1097/MNH.0000000000000609
2. Jahnen-Dechent W., Pasch A. Solving the insoluble: calciprotein particles mediate bulk mineral transport. *Kidney Int*. 2023;103(4):663-665. doi 10.1016/j.kint.2023.01.011
3. Shishkova D., Lobov A., Repkin E. et al. Calciprotein Particles Induce Cellular Compartment-Specific Proteome Alterations in Human Arterial Endothelial Cells. *J Cardiovasc Dev Dis*. 2023;11(1):5. doi 10.3390/jcdd11010005
4. Feenstra L., Kutikhin A.G., Shishkova D.K. et al. Calciprotein Particles Induce Endothelial Dysfunction by Impairing Endothelial Nitric Oxide Metabolism. *Arterioscler Thromb Vasc Biol*. 2023;43(3):443-455. doi 10.1161/ATVBAHA.122.318420
5. Shishkova D., Lobov A., Zainullina B. et al. Calciprotein Particles Cause Physiologically Significant Pro-Inflammatory Response in Endothelial Cells and Systemic Circulation. *Int J Mol Sci*. 2022;23(23):14941. doi 10.3390/ijms232314941
6. Bogdanov L., Shishkova D., Mukhamadiyarov R. et al. Excessive Adventitial and Perivascular Vascularisation Correlates with Vascular Inflammation and Intimal Hyperplasia. *Int J Mol Sci*. 2022;23(20):12156. doi 10.3390/ijms232012156
7. Shishkova D.K., Velikanova E.A., Bogdanov L.A. et al. Calciprotein Particles Link Disturbed Mineral Homeostasis with Cardiovascular Disease by Causing Endothelial Dysfunction and Vascular Inflammation. *Int J Mol Sci*. 2021;22(22):12458. doi 10.3390/ijms222212458

8. Kutikhin A.G., Feenstra L., Kostyunin A.E., Yuzhalin A.E., Hillebrands J.L., Krenning G. Calciprotein Particles: Balancing Mineral Homeostasis and Vascular Pathology. *Arterioscler Thromb Vasc Biol*. 2021;41(5):1607-1624. doi 10.1161/ATVBAHA.120.315697
9. Shishkova D., Markova V., Sinitsky M., Tsepokina A., Velikanova E., Bogdanov L., Glushkova T., Kutikhin A. Calciprotein Particles Cause Endothelial Dysfunction under Flow. *Int J Mol Sci*. 2020;21(22):8802. doi 10.3390/ijms21228802
10. Shishkova D., Velikanova E., Sinitsky M., Tsepokina A., Gruzdeva O., Bogdanov L., Kutikhin A. Calcium Phosphate Bions Cause Intimal Hyperplasia in Intact Aortas of Normolipidemic Rats through Endothelial Injury. *Int J Mol Sci*. 2019;20(22):5728. doi 10.3390/ijms20225728
11. Kutikhin A.G., Velikanova E.A., Mukhamadiyarov R.A. et al. Apoptosis-mediated endothelial toxicity but not direct calcification or functional changes in anti-calcification proteins defines pathogenic effects of calcium phosphate bions. *Sci Rep*. 2016;6:27255. doi 10.1038/srep27255

doi 10.18699/bgrs2024-9.1-24

Identification of *SDHx* variants with gene expression model in head and neck paragangliomas

Snezhkina A.V.[1]*, Kobelyatskaya A.A.[1], Ayupova A.F.[1], Fedorova M.S.[1], Pavlov V.S.[1], Lantsova M.S.[1], Kuzovkina N.S.[1], Pudova E.A.[1], Kalinin D.V.[2], Kudryavtseva A.V.[1]

[1] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*

[2] *Vishnevsky Institute of Surgery, Ministry of Health of the Russian Federation, Moscow, Russia*

* *leftger@rambler.ru*

Key words: Head and neck paragangliomas; *SDHx* variants; gene expression; prediction model

Motivation and Aim: Head and neck paraganglioma (HNPGL) is a rare neuroendocrine neoplasm with a high rate of hereditary predisposition. Both germline and somatic mutations in the *SDHx* genes occur in over than 40 % of HNPGLs. Variants in the *SDHB* and *SDHD* genes are associated with an increased risk of metastatic and multifocal tumor forms, respectively. The objective of the study is to identify mRNA gene expression changes associated with *SDHx* variants and develop an expression model for mutation detection.

Methods and Algorithms: Whole-transcriptome sequencing data obtained on an Illumina platform for HNPGLs were utilized for differential gene expression analysis between *SDHx*-mutated and non-mutated tumors. Based on counts per million (CPM) values, a fully connected neural network (FCNN) was constructed using Keras, Tensorflow, and KerasR libraries. The obtained expression model was trained and tested both on a study cohort and the TCGA-PCPG RNA-Seq data. Subsequently, the model was validated on an extended tumor set (approximately 200 FFPE samples) using quantitative PCR (qPCR). The work was performed using the equipment of the EIMB RAS "Genome" center (http://www.eimb.ru/rus/ckp/ccu_genome_c.php).

Results: A two-step gene expression model for the identification of variants in *SDHx* genes using qPCR has been developed. At the initial step, the model defines variants in *SDHx* genes in general based on the mRNA expression of four genes – *CEP104*, *EYA4*, *FGD5*, and *NFRKB*. The first gene set demonstrated the following metrics in the test: sensitivity – 0.87, specificity – 0.94, accuracy – 0.90, and area under the curve (AUC) – 0.90. In the second step, the model predicts *SDHB* and *SDHD* variants using the expression mRNA data of *CEP104*, *EYA4*, *FGD5*, and *TERC* genes, and the following metrics were observed in the test: sensitivity – 1, specificity – 0.77, accuracy – 0.89, and AUC – 0.87.

Conclusion: The most effective method for identifying variants in the *SDHx* genes is genetic testing with targeted sequencing or whole-exome sequencing. However, these methods are expensive and time consuming. In this study, we created and validated an expression model for the prediction of *SDHx* variants in HNPGLs that can be used in routine clinical practice. In addition, the model enables the identification of *SDHB* and *SDHD* variants, which is very important for HNPGL management.

Funding: The study is supported by a grant from the Russian Science Foundation (No. 24-14-00439).

9 Симпозиум «Биомедицина, биоинформатика и системная компьютерная биология»

Symposium “Biomedicine, bioinformatics and systems computational biology”

doi 10.18699/bgrs2024-9.2-01

Семаглутид способствует усилению термогенеза бежевых адипоцитов, полученных из мезенхимальных стволовых клеток подкожной жировой ткани пациентов с сахарным диабетом 2 типа

Агарёва М.Ю.[1, 2]*, Стафеев Ю.С.[1], Мичурина С.С.[1], Шестакова Е.А[3], Томилова А.О.[3], Синеокая М.С.[3], Меньшиков М.Ю.[1], Парфёнова Е.В.[1, 2], Шестакова М.В.[3]

[1] *Национальный медицинский исследовательский центр кардиологии имени академика Е.И. Чазова Министерства здравоохранения Российской Федерации, Москва, Россия*

[2] *Московский государственный университет имени М.В. Ломоносова, Москва, Россия*

[3] *Национальный медицинский исследовательский центр эндокринологии Министерства здравоохранения Российской Федерации, Москва, Россия*

**amarrgo1999@gmail.com*

Ключевые слова: семаглутид; мезенхимальные стволовые клетки; сахарный диабет 2 типа

Мотивация и цель: В настоящее время агонисты рецептора глюкагоноподобного пептида-1 (ГПП-1) широко используются для терапии сахарного диабета 2 типа (СД2Т). Известно, что инъекции семаглутида способствуют снижению веса и предотвращают развитие осложнений, ассоциированных с СД2Т. Семаглутид оказывает положительное влияние на различные органы и ткани, среди которых и жировая ткань (ЖТ), играющая важную роль в регуляции метаболического статуса. Активация бежевых адипоцитов приводит к повышению чувствительности к инсулину и снижению содержания глюкозы в крови, что предотвращает ожирение и СД2Т. Мезенхимальные стволовые клетки (МСК) являются важным компонентом жировой ткани, они обладают способностью дифференцироваться в адипоциты, обеспечивая самообновление и расширение ткани. При СД2Т дифференцировка МСК нарушается, что приводит к прогрессированию заболевания и развитию сопутствующих осложнений. Цель нашей работы – исследовать влияние агониста рецептора ГПП-1 семаглутида на бежевую адипогенную дифференцировку МСК ЖТ и термогенную активность полученных адипоцитов.

Методы и алгоритмы: МСК были получены из подкожной жировой ткани пациентов с СД2Т до терапии семаглутидом и спустя 6 месяцев после терапии (ИМТ>35 кг/м^2, длительность ожирения более 10 лет). Выделение клеток проводили согласно стандартному протоколу с использованием коллагеназы 1 типа. Дифференцировку в бежевом направлении осуществляли в течение 21 дня в присутствии адипогенных индукторов. Размер липидных капель оценивали с помощью красителя BODIPY493/503, термогенную активность – с помощью термофильного красителя ERThermAC. Экспрессию белков-маркеров адипогенеза, воспаления и бежевой дифференцировки оценивали методом иммуноблоттинга.

Результаты: Мы обнаружили, что после терапии семаглутидом количество бежевых адипоцитов в поле зрения увеличивается, что говорит об усилении

адипогенного потенциала. Также после терапии содержание мелких липидных капель в бежевых адипоцитах выше, чем в бежевых адипоцитах до терапии, кроме того, увеличивается экспрессия HSL, что говорит о более активном липолизе в адипоцитах после терапии. Экспрессия маркера воспаления RAGE снижается после терапии, а экспрессия меркера термогененза UCP1 – увеличивается, как и термогенная активность бежевых адипоцитов, определенная с помощью термочувствительного красителя.
Выводы: Таким образом, терапия агонистом рецептора ГПП-1 семаглутидом способствуют усилению термогенной и липолитической активности бежевых адипоцитов, полученных из мезенхимальных стволовых клеток подкожной жировой ткани пациентов с СД2Т.
Финансирование: Исследование поддержано грантом Российского научного фонда № 22-15-00365.

Semaglutide enhances the thermogenesis of beige adipocytes derived from mesenchymal stem cells of subcutaneous adipose tissue of patients with type 2 diabetes mellitus

Agareva M.[1,2]*, Stafeev I.[1], Michurina S.[1], Shestakova E.[3], Tomilova A.[3], Sineokaya M.[3], Menshikov M.[1], Parfyonova Ye.[1,2], Shestakova M.[3]

[1] *National Medical Research Centre of Cardiology named after academician E.I. Chazov, Moscow, Russia*

[2] *Lomonosov Moscow State University, Moscow, Russia*

[3] *Endocrinology Research Centre, Moscow, Russia*

* *amarrgo1999@gmail.com*

Key words: semaglutide, mesenchymal stem cells, type 2 diabetes mellitus

Motivation and Aim: Currently, glucagon-like peptide-1 (GLP-1) receptor agonists are widely used for the treatment of type 2 diabetes mellitus (T2DM). Semaglutide injections are known to promote weight loss and prevent the development of complications associated with T2DM. Semaglutide has a positive effect on various organs and tissues, including adipose tissue (AT), which plays an important role in the regulation of metabolic status. Activation of beige adipocytes leads to increased insulin sensitivity and decreased blood glucose, which prevents obesity and T2DM. Mesenchymal stem cells (MSC) are an important component of adipose tissue; they have the ability to differentiate into adipocytes, ensuring self-renewal and tissue expansion. In T2DM, MSC differentiation is impaired, which leads to disease progression and the development of associated complications. The purpose of our work is to investigate the effect of the GLP-1 receptor agonist semaglutide on the beige adipogenic differentiation of AT MSCs and the thermogenic activity of the resulting adipocytes.
Methods and Algorithms: MSCs were obtained from the subcutaneous adipose tissue of patients with T2DM before therapy with a GLP-1 receptor agonist, and 6 months after therapy from repeated biopsies of the subcutaneous adipose tissue of the same patients (BMI>35 kg/m^2, duration of obesity more than 10 years). Cells were isolated according to a standard protocol using type 1 collagenase. Differentiation in the beige direction was carried out for 21 days in the presence of adipogenic inducers. The size of lipid

droplets was assessed using the BODIPY493/503 dye, and thermogenic activity was assessed using the thermophilic dye ERThermAC. The expression of protein markers of adipogenesis, inflammation and beige differentiation was assessed by immunoblotting.
Results: We found that after treatment with semaglutide, the content of beige adipocytes in the visual field increases, which indicates an increase in adipogenic potential. Also, after therapy, the content of small lipid droplets in beige adipocytes is higher compared to adipocytes before therapy, as is the expression of HSL, which indicates more active lipolysis in adipocytes after therapy. The expression of the inflammatory marker RAGE decreases after therapy, and the expression of the thermogenic marker UCP1 increases, as does the thermogenic activity of beige adipocytes, determined using a thermophilic dye.
Conclusion: Thus, therapy with the GLP-1 receptor agonist semaglutide enhances the thermogenic and lipolytic activity of beige adipocytes obtained from mesenchymal stem cells of subcutaneous adipose tissue of patients with T2DM.
Funding: The study was supported by the Russian Science Foundation grant No. 22-15-00365.

doi 10.18699/bgrs2024-9.2-02

Аптамеры к белкам, ассоциированным с воспалением – перспективы применения для биомедицины

Воробьева М.А.

Институт химической биологии и медицины СО РАН, Новосибирск, Россия
maria@vorobjeva.ru

Ключевые слова: олигонуклеотидные аптамеры; воспаление; цитокины; тест-системы; ингибирование активности

Мотивация и цель: Олигонуклеотидные аптамеры способны специфично и с высоким сродством связывать заданные молекулы-мишени. В настоящее время аптамеры рассматривают в качестве функциональных аналогов моноклональных антител, при этом по своей природе они являются нуклеиновыми кислотами и поэтому обладают собственными уникальными преимуществами. Наиболее существенные из них – возможность селекции аптамеров «в пробирке», масштабируемый и воспроизводимый химический синтез, точная информация о нуклеотидной последовательности и структуре, стабильность при хранении и транспортировке, возможность введения широкого спектра химических модификаций. Использование аптамеров позволяет существенно расширить набор инструментов для решения фундаментальных и прикладных задач, связанных с молекулярным узнаванием. В частности, очень перспективным, но пока относительно мало разработанным направлением в этой области является создание новых способов диагностики и терапии иммуновоспалительных заболеваний. Распространенность и хронический характер этих патологий, важность их ранней, в том числе доклинической диагностики, обусловливают потребность в доступных тест-системах, пригодных для долговременного мониторинга биомаркеров. Аптамеры могут быть использованы в качестве узнающих элементов для таких тест-систем. С другой стороны, аптамеры, способные к таргетному ингибированию белков-участников воспалительных каскадов, представляют интерес как потенциальные терапевтические средства.

Методы и алгоритмы: Наши исследования направлены на создание нового комплексного подхода к диагностике и терапии анкилозирующего спондилита с использованием ДНК- и РНК-аптамеров, направленных на важные для его патогенеза белки: провоспалительные цитокины фактор некроза опухоли α и интерлейкин-17А, а также ингибитор Wnt-пути белок Dickkopf-1. В рамках этой работы были получены серии аптамеров к выбранным биомаркерам и исследован их потенциал в качестве компонентов колориметрических тест-систем для количественной детекции белков-мишеней. Исследована также возможность использования данных аптамеров и их модифицированных производных для подавления функциональной активности белков-мишеней на моделях хронического и острого воспаления.

Результаты: Получены новые аптамеры к белкам, ассоциированным с воспалением. Созданы колориметрические системы на основе аптамеров к растворимым белкам ФНОα, DKK-1, IL-17A, показана их работоспособность. Показано, что аптамеры к IL-17A и ФНОα способны ингибировать активность белков-мишеней в системах *in vitro* и *in vivo*.

Выводы: Полученные аптамеры являются перспективной основой для создания новых средств диагностики и терапии воспалительных заболеваний.

Финансирование: Исследование выполнено при финансовой поддержке гранта РНФ и правительства НСО № 22-15-20050 и государственного задания ИХБФМ СО РАН № 121031300042-1.

Aptamers to inflammation-associated proteins and their prospects for biomedical applications

Vorobyeva M.A.

Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
maria@vorobjeva.ru

Key words: oligonucleotide aptamers, inflammation, cytokines, test systems, inhibition of activity

Motivation and Aim: Oligonucleotide aptamers are capable of binding given target molecules specifically and with high affinity. Currently, aptamers are considered as functional analogues of monoclonal antibodies, while by their nature they are nucleic acids and therefore have their own unique advantages. The most significant of them are the possibility of selection on a lab bench, scalable and reproducible chemical synthesis, precise information about the nucleotide sequence and structure, stability during storage and transportation, and the possibility of introducing a wide range of chemical modifications. The use of aptamers significantly expands the set of tools for solving fundamental and applied issues related to molecular recognition. In particular, a very promising, but so far relatively poor developed direction in this area is the creation of new means for diagnosing and treating immunoinflammatory diseases. The prevalence and chronic nature of these pathologies, as well as the importance of their early, especially preclinical diagnosis, determine the need for easy handling test systems suitable for long-term monitoring of biomarkers. Aptamers can suit as recognition elements for such test systems. Otherwise, aptamers capable of targeted inhibition of proteins involved in inflammatory cascades are of interest as potential therapeutic agents.

Methods and Algorithms: Our research develops a new complex approach to the diagnosis and treatment of ankylosing spondylitis using DNA and RNA aptamers targeting proteins important for its pathogenesis: pro-inflammatory cytokines tumor necrosis factor α and interleukin-17A, and a Wnt pathway inhibitor Dickkopf-1. We obtained a series of aptamers for selected biomarkers and investigated their potential as components of colorimetric test systems for the quantitative detection of target proteins. The possibility of using these aptamers and their modified derivatives to suppress the functional activity of target proteins in models of chronic and acute inflammation was also studied.

Results: New aptamers have been generated for proteins associated with inflammation. Colorimetric systems based on aptamers for the soluble proteins TNFα, DKK-1, IL-17A have been created, and their workability has been proved. Aptamers to IL-17A and TNFα demonstrated their potential of inhibiting the activity of target proteins *in vitro* and *in vivo*.

Conclusion: The obtained aptamers represent a promising basis for the development of new diagnostic and therapeutic tools for inflammatory diseases.

Funding: The study was fulfilled with the financial support of the Russian Science Foundation and the Government of the Novosibirsk Region No. 22-15-20050 and the Russian state-funded project for ICBFM SB RAS (grant No. 121031300042-1).

doi 10.18699/bgrs2024-9.2-03

Фосфорилгуанидиновые олигонуклеотиды как перспективные инструменты для создания высокочувствительных гомо- и гетерофазных систем выделения и анализа нуклеиновых кислот

Дмитриенко Е.*, Булгакова А., Чубаров А., Ломзов А., Пышный Д.
Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** elena.dmitrienko@niboch.nsc.ru*

Ключевые слова: Нуклеиновые кислоты; фосфорилгуанидиновые олигонуклеотиды; ПЦР; диагностика; зонды; биосенсоры; система детекции НК-биомаркеров

Мотивация и цель: Нуклеиновые кислоты (НК), их аналоги и производные являются молекулярными инструментами для научных молекулярно-биологических исследований, а также используются для решения практических задач биомедицины, биотехнологии и молекулярной диагностики. Выделение нуклеиновых кислот из биологических образцов и их анализ различными методами представляет собой одну из самых значимых проблем глобального здравоохранения, так как крайне необходимо для клинической диагностики и выявления патологий и возбудителей социально значимых инфекций. Однако низкая эффективность и специфичность выделения НК значительно усложняют дальнейший анализ и приводит к ложноотрицательным или ложноположительным результатам. Введение химических модификаций в состав олигонуклеотидов позволяет направленно изменять их структуру, свойства, взаимодействие с биологически активными соединениями. Синтетические аналоги нуклеиновых кислот часто обладают улучшенными физико-химическими, молекулярно-биологическими и биологическими свойствами. Применение модифицированных олигонуклеотидов, конструирование на их основе диагностических и терапевтических систем, а также использование их в различных гетерофазных форматах для эффективного выделения и анализа позволяет решить проблемы современной диагностики НК.
Методы и алгоритмы: Дизайн и синтез набора нативных и модифицированных ДНК, содержащих в заданных положениях одну или несколько, включая полностью модифицированные, фосфорилгуанидиновых (ФГ) модификаций по межнуклеотидной фосфатной группе проведен методами твердофазного амидофосфитного синтеза на автоматическом НК-синтезаторе. Полученные производные выделены, очищены и охарактеризованы методами ВЭЖХ, электрофореза, ESI либо MALDI TOF масс-спектрометрии, а также спектрофотометрическими методами. Изучены свойства ФГ-содержащих зондов в системах аллель-специфического ПЦР-анализа. Иммобилизацию ФГ-содержащих олигонуклеотидов на различные поверхности, включая наночастицы, проводили с использованием введения функциональных групп в состав олигонуклеотидов, бифункциональных реагентов и функционализированных силанов. Полученные результаты использованы для апробации различных вариантов систем гомофазного и гетерофазного анализа НК при разработке биосенсоров.

Результаты: Проведено детальное исследование влияния количества и положения ФГ-модификаций в составе праймеров на эффективность протекания реакции удлинения цепи с помощью Taq ДНК-полимеразы. Был проведен дизайн и синтез более 40 модельных олигонуклеотидов, содержащих ФГ-модифицированные звенья в различном количестве и в разных положениях от 3′-конца. Все синтезированные олигонуклеотиды были исследованы, как в качестве праймеров при ферментативном удлинении в составе дуплексов с комплементарной 30-звенной матрицей с помощью Taq-ДНК-полимеразы, так и в качестве матриц в составе дуплексов с комплементарной 8-звенной последовательностью флуоресцентно-меченного олигонуклеотида. Для ФГ-модифицированных олигонуклеотидов, которые дают максимальный и минимальный выход продукта, как в случае праймера, так и в случае матрицы, показано возможность их применения в качестве праймеров для амплификации гена eGFP в режиме реального времени с использованием нативного обратного праймера. С учетом полученных данных осуществлен дизайн и синтез набора модифицированных праймеров, содержащих все возможные варианты ближайших соседей и мисматчей непосредственно рядом с введением фосфорилгуанидиновой модификации, для систематического исследования выявления однонуклеотидных полиморфизмов в процессе аллель-специфической полимеразной цепной реакции. Показано, что некоторые пары праймеров ингибируют ПЦР в случае наличия мисматча, и при этом наличие ФГ-модификации приводит к еще большему ингибированию, чем в его отсутствии. Проанализирован пул полученных продуктов ПЦР методом высокопроизводительного секвенирования. В случае использования библиотеки праймеров без мисматча в 3’-положении, наличие ФГ-модификации полностью ингибирует образование несовершенных продуктов. Показано, что включение ФГ-модификации в праймеры для ПЦР приводит к усилению различия между ДНК дикого типа и мутированной ДНК, обеспечивая высокую дискриминирующую способность праймеров в аллель-специфической полимеразной цепной реакции в реальном времени (АС-ПЦР). Разработана система гетерофазного гибридизационного анализа НК с использованием ФГ-зондов. Исследованы методы иммобилизации нативных и модифицированных олигонуклеотидов на различные твердотельные поверхности, в том числе кремниевые микропроволоки, и эффективность гетерофазной молекулярной гибридизации. Разработаны подходы к иммобилизации ФГ-зондов на разные типы поверхности. Показано, что использование ФГ-зондов увеличивает эффективность гетерофазной гибридизации нуклеотидных маркеров в низкосолевых (10 мМ фосфатный буфер, pH 7.4) и в бессолевых условиях. Апробированы протоколы пробоподготовки аналита с использованием композитного материала на основе магнитных наночастиц оксида железа и полимера нейлона-6. Разработана система гетерофазного гибридизационного анализа на КНИ-биосенсоре. Показано, что использование карбонилдиимидазола в качестве активирующего агента позволяет с большей чувствительностью выявлять матрицу в сравнении с активирующим агентом 3-глицидоксипропилтриметоксисиланом. При этом выявление протяженной рРНК матрицы оказалось более эффективным по сравнению с короткой синтетической ДНК-матрицей. Продемонстрирована чувствительность КНИ-биосенсора на аттомолярном уровне.

Выводы: Таким образом, получен набор данных, в еще большей степени раскрывающих потенциал использования фосфорилгуанидиновых олигонуклеотидов для разработки современных систем анализа нуклеиновых кислот, и разработаны алгоритмы направленного конструирования высокоспецифичных зондов и праймеров, позволяющих решать наиболее сложные прикладные задачи в области молекулярной диагностики НК.
Финансирование: Исследование выполнено при поддержке Российского научного фонда (№ 22-24-00996) и государственного задания Института химической биологии и фундаментальной медицины СО РАН (№ 121031300042-1).

Phosphorylguanidine derivatives of oligonucleotide as perspective tools for development of highly sensitive systems for nucleic acids diagnostics

Dmitrienko E.*, Bulgakova A., Chubarov A., Lomzov A., Pyshnyi D.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** elena.dmitrienko@niboch.nsc.ru*

Key words: Nucleic acids; phosphoryl guanidine oligonucleotides; PCR diagnostics; probes; biosensors; detection of NA-biomarkers

Motivation and Aim: Nucleic acids (NA), their analogues and derivatives are molecular tools for scientific molecular biological research and are also used to solve practical problems in biomedicine, biotechnology and molecular diagnostics. The isolation of nucleic acids from biological samples and their analysis by various methods represents one of the most significant problems of global health, as it is essential for clinical diagnosis and detection of pathologies and pathogens of socially important infections. However, the low efficiency and specificity of NA isolation significantly complicate further analysis and lead to false negatives or false positives.
Introduction of chemical modifications into the composition of oligonucleotides allows to change their structure, properties, and interaction with biologically active compounds. Synthetic analogues of nucleic acids often have improved physicochemical, molecular-biological and biological properties. The use of modified oligonucleotides, construction of diagnostic and therapeutic systems based on them, as well as their use in various heterophase formats for efficient isolation and analysis allows solving the problems of modern NA diagnostics.
Methods and Algorithms: The design and synthesis of a set of native and modified DNAs containing one or more (including fully modified) phosphorylguanidine (PG) modifications on the inter-nucleotide phosphate group at given positions were carried out by solid-phase amidophosphite synthesis on an automatic NA synthesizer. The obtained derivatives were isolated, purified and characterized by HPLC, electrophoresis, ESI or MALDI TOF mass spectrometry and spectrophotometric methods. The properties of PG-containing probes in allele-specific PCR assay systems were studied. Immobilisation of PG-containing oligonucleotides on various surfaces, including nanoparticles, was carried out using the introduction of functional groups into the oligonucleotide composition, bifunctional reagents and functionalized silanes. The results obtained were used to validate different variants of homophase and heterophase NA analysis systems for the development of biosensors.

Results: The detailed study of the influence of the number and position of PG modifications in the primers on the efficiency of the chain extension reaction using Taq DNA polymerase was carried out. The design and synthesis of more than 40 model oligonucleotides containing PG-modified units in varying quantities and in different positions from the 3' end were carried out. All synthesized oligonucleotides were studied both as primers for enzymatic extension as part of duplexes with a complementary 30-mer template using Taq-DNA polymerase, and as templates in duplexes with a complementary 8-mer sequence of a fluorescently labeled oligonucleotide. For PG-modified oligonucleotides, which give the maximum and minimum product yield, both in the case of a primer and in the case of a template, the possibility of their use as primers for amplification of the eGFP gene in real time using a native reverse primer has been shown. Taking into account the obtained data, a set of modified primers was designed and synthesized, containing all possible variants of nearest neighbors and mismatches directly next to the introduction of phosphorylguanidine modification, for a systematic study of the detection of single nucleotide polymorphisms in the process of allele-specific polymerase chain reaction. It has been shown that some pairs of primers inhibit PCR in the presence of a mismatch, and the presence of the PG modification leads to even greater inhibition than in its absence. The pool of obtained PCR products was analyzed using high-throughput sequencing. When using a primer library without a mismatch in the 3' position, the presence of the PG modification completely inhibits the formation of imperfect products. It has been shown that the inclusion of PG modification in PCR primers leads to increased discrimination between wild-type and mutated DNA, providing high discriminatory ability of primers in allele-specific real-time polymerase chain reaction (AS-PCR). A system for heterophase hybridization analysis of NA using PG probes has been developed. Methods for immobilization of native and modified oligonucleotides on various solid surfaces, including silicon microwires, and the efficiency of heterophasic molecular hybridization were studied. Approaches to the immobilization of PG probes on different types of surfaces have been developed. It has been shown that the use of PG probes increases the efficiency of heterophase hybridization of nucleotide markers in low-salt (10 mM phosphate buffer, pH 7.4) and in salt-free conditions. Protocols for analyte sample preparation using a composite material based on magnetic iron oxide nanoparticles and nylon-6 polymer have been tested. A system for heterophase hybridization analysis on a SOI biosensor has been developed. It has been shown that the use of carbonyldiimidazole as an activating agent makes it possible to detect the matrix with greater sensitivity in comparison with the activating agent 3-glycidoxypropyltrimethoxysilane. At the same time, identifying an extended rRNA template turned out to be more effective compared to a short synthetic DNA template. The sensitivity of the SOI biosensor at the attomolar level has been demonstrated.

Conclusion: Thus, we have obtained a set of data that further reveal the potential of using phosphorylguanidine oligonucleotides for the development of modern nucleic acid analysis systems, and developed algorithms for the directed design of highly specific probes and primers to solve the most complex applied problems in the field of molecular diagnostics of NA.

Funding: The study is supported by the Russian Science Foundation (grant No. 22-24-00996) and the Ministry of Science and Higher Education of the Russian Federation (state registration No. 121031300042-1).

doi 10.18699/bgrs2024-9.2-04

Исследование мутаций в ткани основных видов опухолей щитовидной железы в первичном и метастатическом очагах

Зяблицкая Е.Ю.*, Асанова Э.Р., Зима Д.В., Максимова П.Е.
Крымский федеральный университет им. В.И. Вернадского, Симферополь, Россия
** evgu79@mail.ru*

Ключевые слова: щитовидная железа, полиморфизм генов, злокачественная опухоль, соматические мутации, метастазы

Мотивация и цель. Злокачественные и доброкачественные опухоли щитовидной железы – это гистологически и генетически гетерогенная группа заболеваний, связанных с трансформацией клеток щитовидной железы, имеющих определенную мутационную нагрузку, которая недостаточно исследована в различных популяциях людей [1]. Однако исследование полиморфизмов генов опухолей представляет большой диагностический интерес для подбора терапии и оценки злокачественного потенциала ткани, а также изучения фундаментальных основ канцерогенеза [2]. Цель нашей работы – изучить особенности полиморфизмов генов щитовидной железы у пациентов при доброкачественных и злокачественных опухолях для клинически значимых мутаций в первичном и метастатическом очагах и выявить возможные закономерности.
Методы и алгоритмы. Исследован тканевой материал (стружка с FFPE-блоков) удаленных щитовидных желез 71 пациента в возрасте 20–60 лет (среднее значение 44.2 года) с верифицированными диагнозами: папиллярный рак (ПР) n = 31, фолликулярный вариант папиллярного рака (ФВПР) n = 10, фолликулярная аденома (ФА) n = 16, контрольная группа аутоиммунный тиреоидит (АИТ) n = 14. Для молекулярно-генетической диагностики использовали наборы реагентов «ДНК-Ткань-Ф» (выделение геномной ДНК) и наборы «Тест-BRAF-ткань», «Тест-KRAS-ткань» и «Тест-EGFR-ткань» (определение статуса мутаций в указанных генах) производителя ТестГен, Россия. Математический анализ осуществляли с использованием «MS Excel».
Результаты. При классическом папиллярном варианте ПР в 54 % случаев выявлены мутации искомых генов. В 58 % случаев клетки имеют мутации BRAF V600E, однако в 29 % случаев выявлены мутации в гене BRAF V600K, редко описываемые в литературе, а в 16 % случаев – мутации в гене KRAS (12 и 13 кодон); 35 % пациентов имеют сочетанные варианты мутаций.
Также интерес представляют мутации, выявленные в метастатических очагах папиллярных карцином: 32 % пациентов этой группы имели метастазы, при этом 40% пациентов с метастатическими поражениями не имели мутаций в первичном очаге, и также 40% имели мутации в тканях лимфогенных метастазов. В такой небольшой выборке встречались любые варианты конкордантности и дискордантности, когда в первичном очаге не было мутаций, а в метастазе мутация KRAS G12V, когда мутации были идентичны (чаще в гене BRAF), или отсутствовали в лимфогенных отсевах опухоли.
При ФВПР щитовидной железы мутации BRAF V600E и BRAF V600K обнаружены в 10 и 20 % случаев соответственно. Мутации в гене KRAS выявлены

как у пациентов с ФА (18 %), так и с ФВПК (20 %). Делеция в гене EGFR обнаружена в 1 случае ФВПР. Таким образом, ФВПР продемонстрировал более широкий спектр мутаций (BRAF, KRAS и EGFR), для ФА характерна только KRAS-мутация.
В образцах с АИТ без опухолевой патологии мутаций обнаружено не было.
Выводы: Таким образом, выявленные мутации указывают на закономерности опухолевой прогрессии и эволюционного развития популяций клеток первичного очага и метастазов, указывают на высокую мутационную нагрузку фолликулярных аденом, подчеркивая сложность биологии фолликулярных неоплазий щитовидной железы, указывают на вероятную эффективность терапии таргетными препаратами.
Финансирование: Исследование поддержано FZEG-2023-0009 «Изучение гетерогенности микроокружения опухоли как фактора ее агрессивности и резистентности к терапии», № 123030700011-4 от 07.03.2023.

Study of mutations in the tissue of the main types of thyroid tumors in primary and metastatic foci

Zyablitskaya E.*, Asanova E., Zima D., Maksimova P.
V.I. Vernadsky Crimean Federal University, Simferopol, Russia
** evgu79@mail.ru*

Key words: thyroid gland, gene polymorphism, malignant tumor, somatic mutations, metastases

Motivation and Aim. Malignant and benign thyroid tumors are a histologically and genetically heterogeneous group of diseases associated with the transformation of thyroid cells with a certain mutational load, which has not been sufficiently studied in various human populations [1]. However, the study of tumor gene polymorphisms is of great diagnostic interest for selecting therapy and assessing the malignant potential of tissue, as well as studying the fundamental principles of carcinogenesis [2]. The purpose of our work is to study the features of thyroid gene polymorphisms in patients with benign and malignant tumors for clinically significant mutations in the primary and metastatic foci and to identify possible patterns.
Methods and Algorithm. Tissue material (shavings from FFPE blocks) of removed thyroid glands of 71 patients aged 20–60 years (average value 44.2 years) with verified diagnoses was studied: papillary cancer (PC) $n = 31$, follicular variant of papillary cancer (FVPC) $n = 10$, follicular adenoma (FA) $n = 16$, control group autoimmune thyroiditis (AIT) $n = 14$. For molecular genetic diagnostics, we used the “DNA-Tissue-F” reagent kits (isolation of genomic DNA) and the “Test-BRAF-tissue”, “Test-KRAS-tissue” and “Test-EGFR-tissue” kits (determining the status of mutations in specified genes) manufacturer TestGen, Russia. Mathematical analysis was carried out using MS Excel.
Results. In the classic papillary variant of PC, mutations of the desired genes were identified in 54 % of cases. In 58 % of cases, the cells have BRAF V600E mutations, but in 29 % of cases mutations in the BRAF V600K gene, rarely described in the literature, were detected, and in 16% of cases – mutations in the KRAS gene (codons 12 and 13); 35 % of patients have combined variants of mutations.

Also of interest are the mutations identified in the metastatic foci of papillary carcinomas: 32 % of patients in this group had metastases, while 40% of patients with metastatic lesions did not have mutations in the primary lesion, and also 40% had mutations in the tissues of lymphogenous metastases. In such a small sample, there were any variants of concordance and discordance, when there were no mutations in the primary focus, and the KRAS G12V mutation in the metastasis, when the mutations were identical (usually in the BRAF gene), or were absent in the lymphogenous screenings of the tumor.

In the FVPC, BRAF V600E and BRAF V600K mutations are found in 10 and 20 % of cases, respectively. Mutations in the KRAS gene were identified in both patients with FA (18%) and FVPC (20 %). A deletion in the EGFR gene was found in 1 case of FVPK. Thus, FVPC demonstrated a wider range of mutations (BRAF, KRAS and EGFR), while FA was characterized only by the KRAS mutation. No mutations were found in samples with AIT without tumor pathology.

Conclusion: Thus, the identified mutations indicate patterns of tumor progression and evolutionary development of cell populations of the primary lesion and metastases, indicate a high mutational load of follicular adenomas, emphasizing the complexity of the biology of follicular neoplasia of the thyroid gland, and indicate the likely effectiveness of therapy with targeted drugs.

Funding: The study is supported by FZEG-2023-0009 “Study of the heterogeneity of the tumor microenvironment as a factor in its aggressiveness and resistance to therapy,” No. 123030700011-4 dated 03/07/2023.

Список литературы/References

1. Иванов А.А., Авдалян А.М., Гервальд В.Я. и др. Мутация BRAF V600E при папиллярном раке щитовидной железы, клиникоморфологические параллели и прогноз. *Российский онкологический журнал*. 2017;22(1):15-20. doi 10.18821/1028-9984-2017-22-1-15-20
   [Ivanov A.A., Avdalyan A.M., Gerval’d V.J. et al. The BRAF mutation V600E in papillary thyroid cancer, clinical-morphological paralleles and prognosis. *Rossiiskii Onkologicheskii Zhurnal = Russian Journal of Oncology*. 2017;22(1):15-20. doi 10.18821/1028-9984-2017-22-1-15-20 (in Russian)]
2. Ромащенко П.Н., Майстренко Н.А., Криволапов Д.С., Симонова М.С. Молекулярно-генетические исследования в хирургии щитовидной железы. *Таврический медико-биологический вестник*. 2021;24(2):118-126
   [Romashchenko P.N., Maistrenko N.A., Krivolapov D.S., Simonova M.S. Molecular-genetic testing in thyroid surgery. *Tavricheskii Mediko-biologicheskii vestnik = Tauride Medical and Biological Bulletin*. 2021;24(2):118-126 (in Russian)]

doi 10.18699/bgrs2024-9.2-05

Предикторная значимость полиморфизмов генов VEGFA, eNOS, IFNL3, IL-6, TP53, ITGA2 в диагностике колоректального рака

Зяблицкая Е.Ю.*, Сеферов Б.Д., Головкин И.О.

Крымский федерльный университет им. В.И. Вернадского, Симферополь, Россия

** evgu79@mail.ru*

Ключевые слова: колоректальная карцинома; полиморфизм генов; злокачественная опухоль; соматические мутации; микроокружение опухоли

Мотивация и цель: Колоректальные карциномы (КРК) занимают лидирующие позиции в структуре заболеваемости раком населения в России и мире (6 %). Среди злокачественных новообразований различных локализаций их доля неуклонно растет, что связано как с наследственными, так и с внешними факторами [1]. Исследование полиморфизмов генов злокачественных новообразований представляет ценность для диагностики, таргетной терапии, прогноза рецидивов и оценки злокачественности, изучения фундаментальных основ опухолевого роста [2]. При этом в опухолевом очаге важно оценивать не только клетки карциномы, но и экстрацеллюлярный матрикс, представляющий собой совокупность клеток микросреды опухоли, содействующих или препятствующих ее росту (сосуды и строма, интратуморозный инфильтрат) [3]. Цель нашей работы – оценить предикторную значимость полиморфизмов генов VEGFA, eNOS, IFNL3, IL-6, TP53, ITGA2 КРК.

Методы и алгоритмы: Исследован тканевой материал (стружка с FFPE-блоков) колоректальных карцином пациентов крымской популяции в возрасте 35–72 лет. Изучали ПЦР-методом полиморфизмы генов микроокружения опухоли: фактора роста эндотелия сосудов, эндотелиальной синтазы оксида азота, матриксных металлопротеиназ 9,12, интерлейкина 6, 28, индуцируемого гипоксией фактора 1-альфа, интегрина альфа-2 и p53 у пациентов с верифицированным колоректальным раком и условно здоровых субъектов, неотягощенных анамнезом. Использовали ДНК-амплификатор CFX-96 и праймеры от производителя Синтол. Математический анализ осуществляли с использованием «MS Excel». Для статистической обработки использовался расчет показателя отношения шансов (ОШ) и доверительного интервала (ДИ).

Результаты: Генотипировано 15 различных полиморфизмов, три из которых могут быть связаны с генетической предрасположенностью к развитию КРК. Носительство аллеля С полиморфизма rs2146323 VEGFA оказывает протективный эффект (ОШ: 0.404; ДИ: 0.22–0.71, расчет из частоты встречаемости аллеля), а аллеля А связано с риском КРК (ОШ: 2.4; ДИ: 1.3–4.3), судить о риске или протективности гетерозиготного генотипа судить невозможно. Носительство генотипа GG и AG полиморфизма rs17576 MMP9 оказывает протективный эффект (ОШ: 0.35; ДИ: 0.14–0.85) на развитие КРК. Можно предположить, что помимо связи этого полиморфизма с риском развития рака молочной железы и

колоректального рака он может иметь предрасположенность и к другим видам рака в крымской популяции.
Носительство аллеля C полиморфизма rs1042522 TP53 связано с протективным воздействием (ОШ: 0.23; ДИ: 0.12–0.41), а аллеля G связано с риском (ОШ: 4.3; ДИ: 2.3–7.8).
Выводы: Носительство аллеля C гена VEGF (rs2146323) и генотипа GG и AG гена MMP (rs17576), а также аллеля C гена TP53 (rs1042522) оказывает протективный эффект.
Финансирование: Исследование поддержано FZEG-2023-0009 «Изучение гетерогенности микроокружения опухоли как фактора ее агрессивности и резистентности к терапии», № 123030700011-4 от 07.03.2023.

Predictive significance of polymorphisms of the VEGFA, eNOS, IFNL3, IL-6, TP53, ITGA2 genes in the diagnosis of colorectal cancer

Zyablitskaya E.*, Seferov B.D., Golovkin I.O.
V.I. Vernadsky Crimean Federal University, Simferopol, Russia
** evgu79@mail.ru*

Key words: colorectal carcinoma; gene polymorphism; malignant tumor; somatic mutations; tumor microenvironment

Motivation and Aim: Colorectal carcinomas (CRC) occupy a leading position in the structure of cancer incidence in Russia and in the world (6 %). Among malignant neoplasms of various localizations, their proportion is steadily growing, which is associated with both hereditary and external factors [1]. The study of polymorphisms in genes of malignant neoplasms is valuable for diagnosis, targeted therapy, prognosis of relapses and assessment of malignancy, and study of the fundamental principles of tumor growth [2]. At the same time, in the tumor focus it is important to evaluate not only carcinoma cells, but also the extracellular matrix, which is a set of cells in the tumor microenvironment that promote or prevent its growth (vessels and stroma, intratumoral infiltrate) [3]. The purpose of our work is to evaluate the predictive significance of polymorphisms of the VEGFA, eNOS, IFNL3, IL-6, TP53, ITGA2 genes of CRC.
Methods and Algorithm: Tissue material (shavings from FFPE blocks) of colorectal carcinomas of patients Crimean population aged 35–72 years was studied. Polymorphisms of tumor microenvironment genes were studied using the PCR method: vascular endothelial growth factor, endothelial nitric oxide synthase, matrix metalloproteinases 9,12, interleukin 6, 28, hypoxia-inducible factor 1-alpha, integrin alpha-2 and p53 in patients with verified colorectal cancer and conditionally healthy subjects, not burdened by anamnesis. We used the CFX-96 DNA amplifier and primers from the manufacturer Synthol. Mathematical analysis was carried out using MS Excel. For statistical processing, we used the calculation of odds ratio (OR) and confidence interval (CI).
Results: Fifteen different polymorphisms were genotyped, three of which may be associated with a genetic predisposition to the development of CRC. Carriage of the C allele of the rs2146323 VEGFA polymorphism has a protective effect (OR: 0.404; CI:

0.22–0.71, calculated from the allele frequency), and the A allele is associated with the risk of CRC (OR: 2.4; CI: 1.3–4.3), it is impossible to judge the risk or protectiveness of a heterozygous genotype. Carriage of the GG and AG genotypes of the rs17576 MMP9 polymorphism has a protective effect (OR: 0.35; CI: 0.14–0.85) on the development of CRC. It can be assumed that in addition to the association of this polymorphism with the risk of developing breast and colorectal cancer, it may also have a predisposition to other types of cancer in the Crimean population. Carriage of the C allele of the rs1042522 TP53 polymorphism is associated with a protective effect (OR: 0.23; CI: 0.12–0.41), and the G allele is associated with a risk (OR: 4.3; CI: 2.3–7.8).

Conclusion: Carriage of the C allele of the VEGF gene (rs2146323) and the GG and AG genotypes of the MMP gene (rs17576), as well as the C allele of the TP53 gene (rs1042522) has a protective effect.

Funding: The study is supported by FZEG-2023-0009 “Study of the heterogeneity of the tumor microenvironment as a factor in its aggressiveness and resistance to therapy,” No. 123030700011-4 dated 03/07/2023.

Список литературы/References

1. Состояние онкологической помощи населению России в 2022 году. М.: МНИОИ им. П.А. Герцена, 2022
   [The state of cancer care for the population of Russia in 2022. Moscow: MNIOI im. P.A. Herzen, 2022. (in Russian)] https://oncology-association.ru/wp-content/uploads/2023/08/sop-2022-el.versiya_compressed.pdf 20.03.2024
2. Wang H., Tian T., Zhang J. Tumor-associated macrophages (tams) in colorectal cancer (crc): from mechanism to therapy and prognosis. *Int J Mol Sci*. 2021;22(16):8470
3. Максимова П.Е., Голубинская Е.П., Сеферов Б.Д., Зяблицкая Е.Ю. Колоректальный рак: эпидемиология, канцерогенез, молекулярно-генетические и клеточные механизмы резистентности к терапии (аналитический обзор). *Колопроктология*. 2023;22(2(84)):160-171
   [Maksimova P.E., Golubinskaya E.P., Seferov B.D., Zyablitskaya E.Yu. Colorectal cancer: epidemiology, carcinogenesis, molecular genetic and cellular mechanisms of resistance to therapy (analytical review). *Coloproctology = Coloproctologia*. 2023;22(2(84)):160-171 (in Russian)]

doi 10.18699/bgrs2024-9.2-06

Быстрая и портативная система тестирования для идентификации ДНК *Borrelia burgdorferi*

Ефименко Е.[1]*, Спалвис А.[2], Макашов А.[3]

[1] *Средняя общеобразовательная школа № 225, Санкт-Петербург, Россия*

[2] *Гимназия № 406, Санкт-Петербург, Россия*

[3] *Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия*

** 6888008@bk.ru*

Современные диагностические системы, основанные на обнаружении нуклеиновых кислот, такие как полимеразная цепная реакция (ПЦР) в реальном времени, позволяют сравнительно быстро диагностировать заболевания, вызываемые РНК содержащими вирусами, и обладают высокой чувствительностью. Кроме того, они достаточно просты в разработке, и их производство может быть быстро поставлено на поток при вспышках различных инфекций. Однако методы, основанные на ПЦР, требуют дополнительной стадии обратной транскрипции – синтеза на основе выделенной из образца РНК цепи комплементарной ДНК (кДНК). Кроме того, амплификация нуклеиновых кислот с помощью ПЦР требует многократно повторяющихся циклов нагрева и охлаждения, что значительно увеличивает временные затраты на проведение теста. Также следует отметить, что для самого процесса пробоподготовки – выделения РНК из клинического образца, требуется наличие специального оборудования и реактивов. Кроме того, дополнительные операции повышают риск как ложноотрицательного (потери генетического материала вируса на этапе выделения), так и ложноположительного (контаминация образцов) результатов. В результате требуется проведение достаточно сложных операций, длительное время реакций и относительно дорогое оборудование. Таким образом, в идеальной диагностической системе число шагов должно быть сведено к минимуму, но без ущерба чувствительности.

Известно, что Bst ДНК-полимераза обладает также и ревертазной активностью, что позволяет использовать её для одношаговой детекции РНК содержащих вирусов. Благодаря подобным свойствам Bst-полимеразы, был разработан метод strand exchange amplification (SEA). Однако параллельное использование рестриктаз позволяет значительно увеличить чувствительность данного метода. В данном случае фрагменты нуклеиновых кислот, полученные в результате воздействия рестриктаз, способны служить в качестве праймеров, увеличивая тем самым скорость накопления детектируемых продуктов, что значительно снижает время проведения анализа без снижения чувствительности по сравнению с классическими методами, основанными на ПЦР.

Подобный метод был использован для диагностики вируса Зика (Ma et al., 2018) и продемонстрировал высокую скорость – время анализа заняло 25 минут – и чувствительность – образцы с концентрацией РНК 10^{-18} моль/л давали положительный результат. Ввиду того что метод основан на принципе изотермической амплификации, он не требует специального дорогостоящего оборудования в отличие от методов, основанных на ПЦР в реальном времени.

Также следует отметить, что визуализация результатов производится невооруженным глазом.

Вирус клещевого энцефалита относится к роду Flavivirus, его геном представлен одноцепочечной молекулой РНК (+ssRNA). Согласно данным Всемирной организации здравоохранения (Еженедельный эпидемиологический бюллетень, № 24, 2011, 86, 241-256, 2011) вирус клещевого энцефалита является важной причиной возникновения вирусной инфекции центральной нервной системы в восточных, центральных и северных странах Европы, северном Китае, Монголии и Российской Федерации. Ежегодно фиксируется до 12000 случаев заражения клещевым энцефалитом, хотя это число считается сильно заниженным. Несмотря на наличие эффективных вакцин, уровень вакцинации не достигает близких к 100% значений, в том числе и в сельских регионах, находящихся в очагах распространения заболевания. Поэтому создание экспресс-системы детекции РНК вируса, а также портативной системы для проведения изотермической амплификации с визуальной детекции, может быть достаточно актуально и востребовано.

Болезнь Лайма вызывается бактериями рода *Borrelia* и переносится иксодовыми клещами. На территории России заболевание встречается повсеместно. Боррелиоз выявляется приблизительно в 1.2% случаев укуса клеща. Прогноз при своевременном начале лечения благоприятный, однако заболевание может вызывать поражения суставов, такие, как хронический артрит. Эффективных вакцин против возбудителя болезни Лайма на сегодняшний день не существует.

Мы планируем применить подобный метод SEA для обнаружения геномной РНК вируса клещевого энцефалита, а также ДНК возбудителя болезни Лайма.

План работ:

Создание тест-системы для детекции РНК вируса клещевого энцефалита.

Подбор специфических праймеров для генома вируса клещевого энцефалита.

Подбор рестриктаз. Подбор праймера, содержащего сайт рестрикции.

Отработка условий изотермической амплификации.

Определение чувствительности метода детекции геномной РНК вируса клещевого энцефалита.

Создание тест-системы для детекции ДНК возбудителя клещевого боррелиоза (*Borrelia sp.*).

Подбор специфических праймеров для консервативных участков генома *Borrelia sp.*

Подбор рестриктаз. Подбор праймера, содержащего сайт рестрикции.

Отработка условий изотермической амплификации.

Определение чувствительности метода ДНК возбудителя клещевого боррелиоза.

doi 10.18699/bgrs2024-9.2-07

Количественные изменения проонкогенных и опухоль-супрессирующих микроРНК в брыжеечном лимфатическом узле при экспериментальном раке молочной железы и фотодинамической терапии

Кабаков А.[1]*, Казаков О.[1], Повещенко А.[1], Черкас В.[1], Бодрова Н.[3], Кононучук В.[1, 2], Гуляева Л.[2]

[1] *НИИ клинической и экспериментальной лимфологии – филиал ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *НИИ молекулярной биологии и биофизики – структурное подразделение ФГБНУ ФИЦ ФТМ, Новосибирск, Россия*

[3] *Новосибирский государственный аграрный университет, Новосибирск, Россия*

** kabakov_av85@mail.ru*

Ключевые слова: рак молочной железы, miRNA, брыжеечный лимфатический узел, фотодинамическая терапия

Мотивация и цель: По последним данным о глобальном бремени рака, в 2020 г. было зарегистрировано около 2,26 млн случаев рака молочной железы, это позволяет говорить, что рак молочной железы является наиболее часто диагностируемым раком в мире. На уровень смертности могут влиять множество факторов, в том числе задержка в диагностике и отсутствие эффективной терапии [1]. Несмотря на множество доступных инструментов в терапии РМЖ, по-прежнему необходимы новые технологии для лечения на всех стадиях рака молочной железы. Относительно новый терапевтический метод, который может оказаться золотым стандартом в лечении неопластических опухолей в сравнении с традиционными методами терапии, является фотодинамическая терапия [2]. МикроРНК представляют собой класс небольших одноцепочечных некодирующих РНК, играющих важную регуляторную роль в экспрессии генов. МикроРНК могут использоваться в качестве диагностических или прогностических биомаркеров при раке молочной железы [3, 4]. Однако количественные изменения проонкогенных miRNA (-21, -27a, -221) и опухоль-супрессирующей miRNA-429 в отдаленных от первичного опухолевого очага лимфатических узлах при экспериментальном раке молочной железы после проведения фотодинамической терапии по литературным данным не были исследованы, что является актуальной научной задачей.

Целью исследования стал анализ количественных изменений проонкогенных и опухоль-супрессирующей miRNAs в отдаленном от первичного опухолевого очага лимфатическом узле у крыс-самок Вистар после проведения фотодинамической терапии экспериментального рака молочной железы.

Методы и алгоритмы: Работа выполнена на половозрелых самках Вистар ($n = 60$). На момент индукции РМЖ возраст крыс-самок Вистар составлял 3 месяца, вес животных составлял 250–300 граммов. Животные участвующие в эксперименте были разделены на 3 группы: 1-я – интактные крысы ($n = 20$); 2-я – крысы с раком молочной железы (РМЖ) без лечения ($n = 20$); 3-я – крысы с РМЖ после

проведения фотодинамической терапии (РМЖ+ФДТ) (n = 20). Индуцирование РМЖ выполнялось 5-кратным подкожным введением N-метил-N-нитрозомочевины (Sigma) с интервалом в 7 дней в область 2-й правой молочной железы. Через 24 недели от начала индукции в 3-й группе крыс-самок Вистар проводилась фотодинамическая терапия РМЖ лазерным аппаратом Российского производства «ЛАХТА-МИЛОН» (группа компаний «МИЛОН»). Процедура ФДТ включает введение в организм тропного к опухоли фотосенсибилизатора. Фотосенсибилизатор «Радахлорин» вводился крысам-самкам Вистар однократно внутрибрюшинно (из расчета 1.2 мг/кг) за 3 ч до проведения лазерного облучения опухоли молочной железы. Световое воздействие на опухоль молочной железы проводили излучением с длиной волны 662 нм в дозе 300 Дж/см^2. Из эксперимента животных выводили через 27 недель от начала индукции РМЖ под наркозом (40 мг/кг внутрибрюшинно нембутала; Sigma), что обуславливало прижизненный забор лимфы. Забор органов: ткани молочной железы, опухолевой ткани и брыжеечных лимфатических, выполнялся после выведения животных из эксперимента. После проведения иммуногистохимического и гистологического исследования, верифицирован аналог люминального В-типа РМЖ человека [5]. Для определения количества проонкогенных miRNA (-21, -27a, -221) и опухоль-супрессирующей miRNA-429 в биологических образцах проводили ОТ-ПЦР в реальном времени с использованием реагентов на амплификаторе CFX96 (Bio-Rad Lab., USA), в качестве гена сравнения использовали РНК U87 (все реагенты производства «Вектор-Бест», Российская Федерация). Статистическую обработку данных проводили с использованием пакета программ Statistica 10, меры центральной тенденции и рассеяния описаны медианой (Me), нижним (Q1) и верхним (Q3) квартилями, достоверность различия рассчитывали по U-критерию Манна-Уитни, и принимали при значениях $p < 0.05$.

Результаты: В группе РМЖ без лечения выявлено увеличение количества miRNA-221 в 4.6 раза в сравнении с интактной группой. В группе РМЖ без лечения отмечено снижение количества проонкогенной miRNA-27a в 4.7 раза в сравнении с интактной группой животных. Значимое снижение количества проонкогенной miRNA-27a в группе РМЖ без лечения отмечалось и в лимфе в 2.3 раза (РМЖ без лечения – 0,48 (0.25–0.91) в сравнении с интактными животными – (1.12 (0.56–2.08), p = 0.005). Проонкогенная miRNA-221 количественно увеличилась в группе РМЖ без лечения в 5.4 раза в сравнении с интактной группой. В брыжеечном лимфатическом узле количество опухоль-супрессирующей miRNA-429 в группе РМЖ без лечения было снижено в 23.6 раза в сравнении с интактными крысами-самками Вистар. Количество miRNA-21 в группе РМЖ без лечения значимо не изменялось. В брыжеечном лимфатическом узле в группе РМЖ после проведения фотодинамической терапии количество проонкогенной miRNA-21 снизилось в 4.5 раза в сравнении с интактной группой. Количество проонкогенной miRNA-27a значимо не изменялось и оставалось на уровне группы РМЖ без лечения. После проведение ФДТ в брыжеечном лимфатическом узле количество проонкогенной miRNA-221 уменьшилось в 5 раз в сравнении с группой РМЖ без лечения. В брыжеечном лимфатическом узле в группе с фотодинамической терапией количество опухоль-супрессирующей miRNA-429 оставалось сниженным в 16.8 раза в сравнении с интактной группой и имела незначительные количественные изменения в сравнении с группой РМЖ без лечения.

Выводы: таким образом, нами установлено снижение проонкогенных микроРНК (miRNA-21 и miRNA-221) в брыжеечном лимфатическом узле после проведения ФДТ РМЖ. При этом количество проонкогенной miRNA-27a сохранялось сниженным и не отличалось от такового в группе РМЖ без лечения. Количество miRNA-429 в брыжеечном лимфатическом узле после ФДТ также оставалось на уровне группы крыс с РМЖ без лечения.

Quantitative changes in pro-oncogenic and tumor-suppressive microRNAs in the mesenteric lymph node in experimental breast cancer and photodynamic therapy

Kabakov A.[1]*, Kazakov O.[1], Poveshchenko A.[1], Cherkas V.[1], Bodrova N.[1, 3], Kononchuk V.[1, 2], Gulyaeva L.[2]

[1] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Research Institute of Molecular Biology and Biophysics – structural subdivision of federal state budgetary scientific institution Federal Research Center for Fundamental and Translational Medicine, Novosibirsk, Russia*

[3] *Federal State Budgetary Educational Institution of Higher Education Novosibirsk State Agrarian University, Novosibirsk, Russia*

** kabakov_av85@mail.ru*

Key words: breast cancer, miRNA, mesenteric lymph node, photodynamic therapy

Motivation and Aim: According to the latest data on the global burden of cancer, there were approximately 2.26 million cases of breast cancer in 2020, making breast cancer the most commonly diagnosed cancer in the world. Moreover, breast cancer is the leading cause of death from tumor diseases among women around the world. Mortality rates can be influenced by many factors, including delay in diagnosis and lack of effective therapy [1]. Despite the many available tools in breast cancer therapy, new technologies are still needed to treat all stages of breast cancer. A relatively new therapeutic method that may prove to be the gold standard in the treatment of neoplastic tumors in comparison with traditional methods of therapy is photodynamic therapy (PDT) [2]. MicroRNAs are a class of small, single-stranded, noncoding RNAs that play important regulatory roles in gene expression. MicroRNAs can be used as diagnostic or prognostic biomarkers in breast cancer [3, 4]. However, quantitative changes in pro-oncogenic miRNA (-21, -27a, -221) and tumor-suppressive miRNA-429 in lymph nodes distant from the primary tumor site in experimental breast cancer after photodynamic therapy and subsequent surgical treatment have not been studied according to the literature, which is an urgent scientific task.

The aim of the study was to analyze quantitative changes in pro-oncogenic and tumor-suppressive miRNAs in a lymph node distant from the primary tumor site in female Wistar rats after photodynamic therapy for experimental breast cancer.

Methods and Algorithms: The work was performed on sexually mature Wistar females (n=60). At the time of breast cancer induction, the age of female Wistar rats was 3 months, the weight of the animals was 250–300 grams. The animals participating in the experiment were divided into 3 groups: 1-st – intact rats (n = 20); 2-nd – rats with

breast cancer (BC) without treatment (n = 20); 3-rd – rats with breast cancer after photodynamic therapy (breast cancer + PDT) (n = 20). Induction of breast cancer was performed by 5-fold subcutaneous injection of N-methyl-N-nitrosourea (Sigma) with an interval of 7 days into the area of the second right breast. After 24 weeks from the start of induction, photodynamic therapy of breast cancer was carried out in the 3rd group of female Wistar rats using a Russian-made laser device "LAKHTA-MILON" (MILON group of companies). The PDT procedure involves the introduction of a tumor-tropic photosensitizer into the body. The photosensitizer "Radachlorin" was administered to female Wistar rats once intraperitoneally (at a rate of 1.2 mg/kg) 3 hours before laser irradiation of the mammary tumor. Light exposure to the breast tumor was carried out with radiation with a wavelength of 662 nm at a dose of 300 J/cm^2. Animals were removed from the experiment 27 weeks from the start of breast cancer induction under anesthesia (40 mg/kg intraperitoneal Nembutal; Sigma), which determined intravital lymph collection. Organ sampling: mammary gland tissue, tumor tissue and mesenteric lymphatic tissue was performed after the animals were removed from the experiment. After immunohistochemical and histological studies, an analogue of human luminal B-type breast cancer was verified [4]. To determine the amount of pro-oncogenic miRNA (-21, -27a, -221) and tumor-suppressive miRNA-429 in biological samples, real-time RT-PCR was performed using reagents on a CFX96 amplifier (Bio-Rad Lab., USA), as U87 RNA was used for the reference gene (all reagents produced by Vector-Best, Russian Federation). Statistical processing of the data was carried out using the Statistica 10 software package, measures of central tendency and dispersion are described by the median (Me), lower (Q1) and upper (Q3) quartiles, the significance of the difference was calculated using the Mann-Whitney U test and was accepted at p values < 0.05.

Results: In the breast cancer group without treatment, an increase in the quantity of miRNA-221 was detected by 4.6-fold compared to the intact group. In the breast cancer group without treatment, there was a 4.7-fold decrease in the quantity of pro-oncogenic miRNA-27a compared to the intact group of animals. A significant decrease in the quantity of pro-oncogenic miRNA-27a in the group of breast cancer without treatment was also observed in the lymph by 2.3-fold (breast cancer without treatment – 0.48 (0.25–0.91) in comparison with intact animals – (1.12 (0.56–2.08), p = 0.005). Pro-oncogenic miRNA-221 increased quantitatively in the breast cancer group without treatment by 5.4 -fold compared to the intact group. In the mesenteric lymph node, the quantity of tumor-suppressive miRNA-429 in the breast cancer group without treatment was reduced by 23.6-fold compared to intact female Wistar rats. The quantity of miRNA-21 in the breast cancer group without treatment did not significantly change. In the mesenteric lymph node in the breast cancer group after photodynamic therapy, the quantity of pro-oncogenic miRNA-21 decreased by 4.5-fold compared to the intact group. The quantity of pro-oncogenic miRNA-27a did not significantly change and remained at the level of the breast cancer group without treatment. After PDT in the mesenteric lymph node, the quantity of pro-oncogenic miRNA-221 decreased by 5 times compared to the breast cancer group without treatment. In the mesenteric lymph node in the group with photodynamic therapy, the quantity of tumor-suppressive miRNA-429 remained reduced by 16.8-fold compared to the intact group and had minor quantitative changes compared to the breast cancer group without treatment.

Conclusion: Hereby, the study revealed a decrease in pro-oncogenic microRNAs (miR-21 and miR-221) in the mesenteric lymph node after photodynamic therapy of BC. At

the same time, the quantity of pro-oncogenic miR-27a in this group remained decreased and did not differ from the group with breast cancer without treatment. The quantity of miR-429 in the regionary lymph node after PDT of breast cancer remained at the level of the group of rats with breast cancer without treatment.

Список литературы/References

1. Siegel R.L. et al. Cancer Statistics, 2021. *CA Cancer J Clin.* 2021;71(1):7-33. doi 10.3322/caac.21654
2. Li J. et al. Non-invasive biomarkers for early detection of breast cancer. *Cancers (Basel)*. 2020;12(10):2767. doi 10.3390/cancers12102767
3. Liang Y.K. et al. MiR-221/222 promote epithelial-mesenchymal transition by targeting Notch3 in breast cancer cell lines. *NPJ Breast Cancer*. 2018;4:20. doi 10.1038/s41523-018-0073-7
4. Казаков О.В., Кабаков А.В., Повещенко А.Ф., Конончук В.В., Струнки Д.Н., Гуляева Л.Ф., Коненков В.И. Влияние фотодинамической терапии на уровень микроРНК в тканях рака молочной железы крыс-самок Вистар. *Бюллетень экспериментальной биологии и медицины*. 2022;173(4):452-455
   [Kazakov O.V., Kabakov A.V., Poveshchenko A.F., Kononchuk V.V., Strunkin D.N., Gulyaeva L.F., Konenkov V.I. Effect of photodynamic therapy on the microRNA level in breast cancer tissues of female Wistar rats. *Bulletin of Experimental Biology and Medicine*. 2022;173(4):444-447]
5. Кабаков А.В., Лыков А.П., Морозов Д.В., Казаков О.В., Повещенко А.Ф., Райтер Т.В., Струнки Д.Н., Коненков В.И. Фенотипическая характеристика химически индуцированной опухоли молочной железы. *Бюллетень экспериментальной биологии и медицины*. 2017;163(4):490-493
   [Kabakov A.V., Lykov A.P., Morozov D.V., Kazakov O.V., Poveshchenko A.F., Raiter T.V., Strunkin D.N., Konenkov V.I. Phenotypical Characteristics of Chemically Induced Mammary Tumor. *Bulletin of Experimental Biology and Medicine.* 2017;163(4):490-492. doi 10.1007/s10517-017-3835-6]

doi 10.18699/bgrs2024-9.2-08

Влияние РМЖ и оперативного лечения РМЖ на количество miR-21, miR-27a, miR-221 и miR-429 в тимусе

Казаков О.[1]*, Кабаков А.[1], Повещенко А.[1], Черкас В.[1], Бодрова Н.[1, 3], Кончук В.[2], Гуляева Л.[2]

[1] *НИИ клинической и экспериментальной лимфологии – филиал ФГБНУ ФИЦ Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *НИИ молекулярной биологии и биофизики – структурное подразделение ФИЦ ФТМ, Новосибирск, Россия*

[3] *Новосибирский государственный аграрный университет, Новосибирск, Россия*

** kazakoff_oleg@mail.ru*

Ключевые слова: рак молочной железы; тимус, микроРНК, оперативное лечение

Мотивация и цель: Изучение взаимосвязей количества микроРНК в тимусе, как центральном органе иммунной системы, при РМЖ и его лечении может иметь значение для понимания участия микроРНК в регуляции иммунного ответа, воздействии на процессы созревания, пролиферации, дифференцировки и активации клеток иммунной системы. Несмотря на то, что уровни онкогенных miR-21, miR-27a, miR-221 и опухоль-супрессирующей miR-429 экспрессируется во многих системах органов, включая иммунную систему, роль этих микроРНК в иммунном ответе организма на РМЖ и оперативное лечение РМЖ остается мало изученной. Цель исследования – сравнить уровни экспрессии микроРНК (miR-21, miR-27a, miR-221, miR-429) в тимусе при РМЖ и после оперативного лечения РМЖ.

Методы и алгоритмы: Работа выполнена на половозрелых самках крыс Вистар (n = 60) с соблюдением принципов надлежащей лабораторной практики и Директивы 2010/63/EU Европейского парламента и Совета ЕС по охране животных, используемых в научных целях. Исследование одобрено локальным этическим комитетом (протокол № 180 от 28.04.2023). Возраст крыс на начало эксперимента — 3 мес (масса 250–300 г). Было сформировано 3 группы животных: 1-я — контрольная (интактные крысы, n=20), 2-я — РМЖ без лечения (n=20), 3-я — оперативное лечение РМЖ (n=20). Из эксперимента животных выводили через 6 мес и 3 нед от момента начала индукции РМЖ под наркозом (нембутал, 40 мг/кг внутрибрюшинно; Sigma). РМЖ моделировали 5-кратным подкожным введением N-метил-N-нитрозомочевины (Sigma) с интервалом 7 сут в область 2-й молочной железы справа. Оперативное лечение РМЖ проводили через 6 мес от момента начала индукции РМЖ, операция под наркозом (нембутал, «Sigma», 40 мг/кг внутрибрюшинно). Забор у животных тимуса выполнялся после выведения их из эксперимента. Тотальную РНК выделяли из ткани вилочковой железы с использованием набора реагентов «Вектор-Бест» по инструкции производителя. Для получения кДНК проводили обратную транскрипцию (ОТ) по матрице микроРНК. Для определения уровней проонкогенных miR-21, miR-27a, miR-221 и опухоль-супрессирующей miR-429 в биологических образцах проводили ОТ-ПЦР в реальном времени на амплификаторе CFX96 (Bio-Rad Lab), в качестве гена

сравнения использовали малую РНК U6 («Вектор-Бест»). Статистическую обработку полученных результатов проводили в программе Statistica 10.0. Меры центральной тенденции и рассеяния описаны медианой (Me), нижним (Q1) и верхним (Q3) квартилем; достоверность различий рассчитывали по *U* критерию Манна–Уитни при $p < 0.05$.

Результаты: В группе РМЖ без лечения, по сравнению с интактной группой, в тимусе уменьшено количество miR-21 в 7.3 раза (контроль – 3.89 (1.0–6.46), РМЖ без лечения – 0.53 (0.37-0.83)), уменьшено количество miR-27a в 10 раз (контроль – 2.0 (0.11–5.3), РМЖ без лечения - 0.19 (0.05-0.32)), уменьшено количество miR-221 в 4.3 раза (контроль – 1.52 (0.5–5.48), РМЖ без лечения – 0.35 (0.09–0.85)). Количество опухоль-супрессирующей miR-429 в тимусе не имеет значимого отличия от показателей интактной группы. Поскольку большинство мишеней miR-21являются супрессорами опухолей, miR-21связана с широким спектром раковых заболеваний, включая РМЖ. MiR-21, известная своей онкогенной активностью, также необходима для реализации иммунных ответов против опухоли [1]. По данным литературы, нокаут miR-21 у мышей замедлял пролиферацию как $CD4^+$, так и $CD8^+$ клеток, снижал выработку ими цитокинов и ускорял рост привитой опухоли. MiR-27a может действовать как супрессор при РМЖ [2]. MiR-27a может нацеливаться на TMEM170B (трансмембранный белок 170B, который участвует в негативной регуляции канонического сигнального пути Wnt), подавлять его, ингибируя сигнальный путь Wnt/β-катенина и подавляя пролиферацию РМЖ [3]. Уменьшение уровня miR-221 в тимусе при РМЖ может быть обусловлено действием miR-221, которая связываясь с p53-зависимым модулятором апоптоза (PUMA), блокирует индуцированный белком p53 апоптоз, влияя на трансформацию клетки и активацию клеточной пролиферации. После оперативного лечения РМЖ, по сравнению с РМЖ без лечения, увеличение в тимусе количества miR-21 (в 3 раза) и miR-27a (в 3.9 раза) может быть обусловлено удалением основного очага опухоли молочной железы. Стрессовые реакции и хирургические манипуляции, вызванные хирургическим вмешательством, могут усиливать метастазирование опухоли за счет высвобождения ангиогенных факторов и подавления естественных клеток-киллеров (NK) и клеточного иммунитета. По данным литературы miR-27a действует как промотор ангиогенеза. По данным литературы miR-21 может активировать $CD4^+$ и $CD8^+$ Т-клетки через сигнальный путь PTEN/Akt, что указывает на ключевые функции miR-21 в обеспечении противоопухолевого иммунного ответа. Уровни экспрессии miR-21 на разных стадиях развития Т-клеток указывают на регуляторную функцию в этом процессе: заметно экспрессируются в незрелых тимоцитах (DN), а в более зрелых клетках снижаются [4]. Mir-21 действует как ключевой регулятор развития и функционирования регуляторных Т-лимфоцитов (Treg) [5], регулирует пролиферацию Treg и экспрессию FOXP3 – транскрипционный фактор, регулирующий транскрипцию генов, ответственных за дифференцировку Т-клеток и экспрессию цитокинов и других факторов, участвующих в супрессии иммунного ответа. Повышенная экспрессия miR-27 в Т-клетках тимуса может отрицательно влиять на индукцию FOXP3 и развитие Treg в тимусе.

Выводы: При РМЖ снижение количества микроРНК (miR-21, miR27a, miR-221) в тимусе может быть обусловлено иммунной реакцией на рост опухоли и акцидентальной инволюцией тимуса. Удаление основного очага опухоли

молочной железы, по сравнению с РМЖ без лечения, вызывает увеличение в тимусе количества miR-21 и miR-27a. Экспрессия опухоль-супрессирующей miR-429 как в группе с РМЖ без лечения, так и после оперативного лечения РМЖ остается на уровне показателей интактной группы.

Influence of breast cancer and surgical treatment of breast cancer on the amount of miR-21, miR-27a, miR-221 and miR-429 in the thymus

Kazakov O.[1]*, Kabakov A.[1], Poveshchenko A.[1], Cherkas V.[1], Bodrova N.[1, 3], Kononchuk V.[2], Gulyaeva L.[2]

[1] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Research Institute of Molecular Biology and Biophysics – structural subdivision of federal state budgetary scientific institution Federal Research Center for Fundamental and Translational Medicine, Novosibirsk, Russia*

[3] *Federal State Budgetary Educational Institution of Higher Education Novosibirsk State Agrarian University, Novosibirsk, Russia*

** kazakoff_oleg@mail.ru*

Key words: mammary cancer; thymus; microRNA; surgical treatment

Motivation and Aim: In breast cancer and its treatment, studying the relationship between the number of microRNAs in the thymus, as the central organ of the immune system, can be important for understanding the participation of microRNAs in the regulation of the immune response, the influence of microRNAs on the processes of maturation, proliferation, differentiation and activation of cells of the immune system. Despite the fact that the levels of oncogenic miR-21, miR-27a, miR-221 and tumor-suppressive miR-429 are expressed in many organ systems, including the immune system, the role of these microRNAs in the body's immune response on the breast cancer and surgical treatment of breast cancer remains little studied. The purpose of the study was to compare the expression levels of microRNAs (miR-21, miR-27a, miR-221, miR-429) in the thymus of breast cancer and after surgical treatment of breast cancer.

Methods and Algorithms: This work was carried out on mature female Wistar rats ($n = 60$), in compliance with the principles of good laboratory practice and Directive 2010/63/EU of the European Parliament and the Council on the protection of animals used for scientific purposes. The local ethics committee (protocol No. 180 of April 28, 2023) approved this study. The age of the rats at the beginning of the experiment was 3 months (weight 250–300 g). 3 groups of animals were formed: 1st – control (intact rats, $n = 20$), 2nd – breast cancer without treatment ($n = 20$), 3rd – surgical treatment of breast cancer ($n = 20$). The animals were removed from the experiment after 6 months and 3 weeks from the start of breast cancer induction under anesthesia (Nembutal, 40 mg/kg intraperitoneally; Sigma). Breast cancer was modeled by 5-fold subcutaneous injection of N-methyl-N-nitrosourea (Sigma) with an interval of 7 days into the area of the 2nd breast on the right. Surgical treatment of breast cancer was carried out 6 months from the start of breast cancer induction, surgery under anesthesia (Nembutal, Sigma, 40 mg/kg intraperitoneally). The thymus was took away from animals after they were

removed from the experiment. Total RNA was isolated from thymus tissue using the Vector-Best reagent kit according to the manufacturer's instructions. To obtain cDNA, reverse transcription (RT) was performed from the microRNA template. To determine the levels of pro-oncogenic miR-21, miR-27a, miR-221 and tumor-suppressive miR-429 in biological samples, real-time RT-PCR was performed on a CFX96 amplifier (Bio-Rad Lab), small RNA U6 was used as a reference gene ("Vector-Best"). Statistical processing of the results was carried out using the Statistica 10.0 program. Measures of central tendency and dispersion are described by median (Me), lower (Q1) and upper (Q3) quartiles; The significance of differences was calculated using the Mann–Whitney U test at $p < 0.05$.

Results: In the breast cancer group without treatment, compared with the intact group, the amount of miR-21 in the thymus was reduced by 7.3 times (control – 3.89 (1.0–6.46), breast cancer without treatment – 0.53 (0.37–0.83)), the amount of miR-27a in 10 times (control – 2.0 (0.11–5.3), breast cancer without treatment – 0.19 (0.05–0.32)), the amount of miR-221 was reduced by 4.3 times (control – 1.52 (0.5–5.48), breast cancer without treatment – 0.35 (0.09–0.85)). The amount of tumor-suppressive miR-429 in the thymus does not differ significantly from the indicators of the intact group. Because most miR-21 targets are tumor suppressors, miR-21 is associated with a wide range of cancers, including breast cancer. MiR-21, known for its oncogenic activity, is also required for the implementation of immune responses against tumors [1]. Knockout of miR-21 in mice slowed down the proliferation of both CD4+ and CD8+ cells, reduced their production of cytokines, and accelerated the growth of the grafted tumor, according to the literature. MiR-27a may act as a suppressor in breast cancer [2]. MiR-27a can target TMEM170B (transmembrane protein 170B, which is involved in the negative regulation of the canonical Wnt signaling pathway) suppress it, inhibiting the Wnt/β-catenin signaling pathway and suppressing breast cancer proliferation [3]. A decrease in the level of miR-221 in the thymus in breast cancer may be due to the action of miR-221, which, by binding to the p53-dependent modulator of apoptosis (PUMA), blocks p53 protein-induced apoptosis, affecting on transformation cell and activation of cell proliferation. After surgical treatment of breast cancer, compared with breast cancer without treatment, removal of the main focus of the breast tumor may be condition an increase in the amount of miR-21 (3 times) and miR-27a (3.9 times) in the thymus. Stress responses and surgical manipulations induced by surgery may enhance tumor metastasis by releasing angiogenic factors and suppressing natural killer (NK) cells and cellular immunity. MiR-27a acts as a promoter of angiogenesis according to the literature. MiR-21 can activate CD4+ and CD8+ T cells through the PTEN/Akt signaling pathway, indicating the key functions of miR-21 in mediating the antitumor immune response. The expression levels of miR-21 at different stages of T cell development indicate a regulatory function in this process: are noticeably expressed in immature thymocytes (DN), and in more mature cells decreases [4]. Mir-21 acts as a key regulator of the development and functioning of regulatory T lymphocytes (Treg) [5], regulates the proliferation of Treg and the expression of FOXP3 – a transcription factor that regulates the transcription of genes responsible for the differentiation of T cells and the expression of cytokines and other factors involved in suppression of the immune response. Increased expression of miR-27 in thymic T cells may negatively influence on induction FOXP3 and development Treg in the thymus.

Conclusion: In breast cancer, a decrease in the amount of microRNAs (miR-21, miR27a, miR-221) in the thymus may be due to an immune response to tumor growth and

accidental involution of the thymus. Removal of the main focus of the breast tumor, compared to breast cancer without treatment, causes an increase in the amount of miR-21 and miR-27a in the thymus. The expression of tumor-suppressive miR-429both in the group with breast cancer without treatment and after surgical treatment of breast cancer remains at the level of the intact group.

Список литературы/References

1. Ward L.D., Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res*. 2016;44(D1):D877-D881. doi 10.1093/nar/gkv1340
2. He W., Wang C., Mu R. et al. MiR-21 is required for anti-tumor immune response in mice: an implication for its bi-directional roles. *Oncogene*. 2017;36(29):4212-4223. doi 10.1038/onc.2017.62
3. Li X., Xu M., Ding L., Tang J. MiR-27a: A Novel Biomarker and Potential Therapeutic Target in Tumors. *J Cancer*. 2019;10(12):2836-2848. doi 10.7150/jca.31361
4. Li M., Han Y., Zhou H. et al. Transmembrane protein 170b is a novel breast tumorigenesis suppressor gene that inhibits the Wnt/β-catenin pathway. *Cell Death Disease*. 2018;9(2):91. doi 10.1038/s41419-017-0128-y
5. Kunze-Schumacher H., Winter S.J., Imelmann E., Krueger A. miRNA miR-21 Is Largely Dispensable for Intrathymic T-Cell Development. *Front Immunol*. 2018;9:2497. doi 10.3389/fimmu.2018.02497
6. Hippen K.L., Loschi M., Nicholls J., MacDonald K.A., Blazar B.R. Effects of MicroRNA on Regulatory T Cells and Implications for Adoptive Cellular Therapy to Ameliorate Graft-versus-Host Disease. *Front Immunol*. 2018;9:57. doi 10.3389/fimmu.2018.00057

doi 10.18699/bgrs2024-9.2-09

Выявление общих и уникальных генетических вариантов, ассоциированных с сахарным диабетом 2 типа и гестационным диабетом

Карпова Н.С.*, Дмитренко О.П., Нурбеков М.К.

Научно-исследовательский институт общей патологии и патофизиологии, Москва, Россия
** nataliiakarpova.sp@gmail.com*

Ключевые слова: полиморфизмы; GWAS; диабет; гестационный сахарный диабет

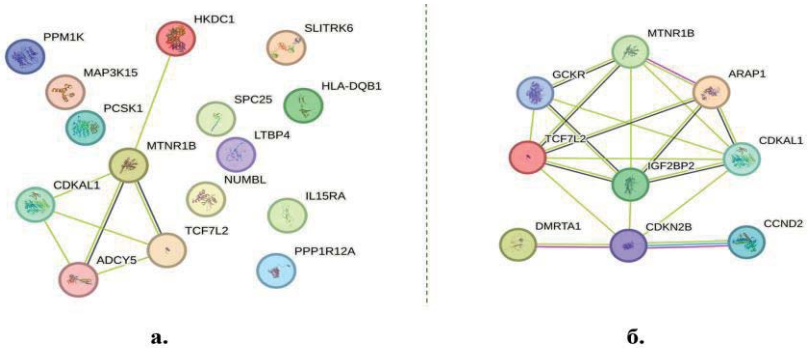
Мотивация и цель: Гестационный сахарный диабет (ГСД) является наиболее часто встречающейся экстрагенитальной патологией беременности и представляет серьезную медицинскую и социальную проблему, увеличивая частоту нежелательных исходов беременности как для матери, так и для плода [1]. В этиопатогенезе гестационного сахарного диабета выявлен генетический фактор. Poulsen с коллегами (2005) выяснили, что генетические факторы объясняют более 75 % нарушения функции β-клеток поджелудочной железы и гомеостаза глюкозы и 53 % развития инсулинорезистентности [2]. Различные стратегии генетического анализа, такие как общегеномный поиск ассоциации (GWAS), применяются для выявления генетических вариантов (включая однонуклеотидные полиморфизмы (SNP)), связанных с ГСД [3–5]. На сегодняшний день идентифицировано около 40 генетических локусов, ассоциированных с ГСД. Часть из них связана с синтезом и секрецией инсулина (INS, KCNJ11, ABCC8, TCF7L2, ND1), с передачей сигнала инсулина (INSR, IGF2, IRS1) и регулированием углеводного и липидного обмена (PPARG, PPARGC1A, ADRB3, GLUT1, ADIPOQ, FOXC2) [3]. Предыдущие исследования показали, что некоторые генетические локусы, которые предрасполагают к развитию сахарного диабета 2 типа (СД2), также могут предрасполагать к ГСД [6]. Поэтому целью данного исследования был поиск общих и уникальных генетических вариантов, связанных с СД2 и ГСД с последующим анализом отобранных полиморфизмов для обнаружения ключевых генов и путей, вовлеченных в развитие ГСД и СД2.
Методы и алгоритмы: Информацию о полиморфизмах, ассоциированных с ГСД и СД2 получали из базы данных GWAS-catalog (https://www.ebi.ac.uk/gwas/home) [3]. Неравновесность сцепления между полиморфизмами оценивали с помощью LDlink (Version 5.6.5, GRch38). Расположение полиморфизмов относительность области генов осуществляли в Ensembl Variant Effect Predictor (VEP) [7]. Анализ белок-белковых взаимодействий (PPI) проводили с помощью STRING DB (Версия 12.0, https://string-db.org) [8]. Из анализа были исключены гены, не кодирующие белки.
Результаты: С 2019 г. были проведены 10 GWAS, в рамках которых обнаружено 30 ассоциаций с ГСД. Из данных полиморфизмов 28 выявлены в рамках одного исследования, в то время как для полиморфизма rs10830963 в гене MTNR1B обнаружена ассоциация с ГСД в нескольких независимых исследованиях (PMID: 36553520, 38182742, 35220425). Согласно LDlink (Version 5.6.5, GRch38) rs10830962 C аллель коррелирует с аллелем C rs10830963 при анализе всех

популяций ($p < 0.0001$; D' = 0.9532; R2 = 0.2718), что свидетельствует в пользу более перспективной для дальнейших исследований ассоциации данного геномного локуса с ГСД.

Мы проанализировали расположение 28 полиморфизмов по отношению к генетическим регионам с помощью Ensembl Variant Effect Predictor (VEP). Большинство полиморфизмов располагается в интронной области гена (67.9 %). Остальные полиморфизмы находятся в межгенной области (10.7 %), области гена (миссенс варианты; 7.1 %), регуляторной области (7.1 %), некодирующего варианта экзона транскрипта (3.6 %), область не определена (3.6 %). Расположение полиморфизмов в регуляторной зоне может означать их влияние на экспрессию генов, образование сплайсинг вариантов генов, а также функционирование регуляторных РНК. Выявлено, что 13 полиморфизмов были ассоциированы как с ГСД, так и с СД2: rs34872471, rs7903146 интрона гена TCF7L2, rs780094 (интрон), rs1260326 (миссенс полиморфизм), rs780093 (интрон) гена GCKR, rs1470579, rs7633675 гена IGF2BP2, rs9348441, rs7754840 интрона гена CDKAL1, rs76895963 в некодирующем экзоне транскрипта CCND2-AS1 и CCND2, rs10811662 в межгенной области CDKN2B-AS1-DMRTA1, rs10830963 интрона гена MTNR1B, rs77464186 интрона гена ARAP1.

Ряд полиморфизмов ассоциирован только с ГСД: rs6798189 гена ADCY5, rs10220124 в межгенной области SLITRK6 – MOB1AP1, rs34499031 интрона гена CDKAL1, rs1402837 интрона гена SPC25, rs1820176 интрона гена MIR583HG и PCSK1, rs6821589 интрон гена PPM1K, rs36090025 интрон гена TCF7L2, rs7254268 в регуляторной области LTBP4-NUMBL, rs10830962 в межгенной области SNRPGP16-MTNR1B, rs9663238 интрона гена HKDC1, rs2274034 интрон гена IL15RA, rs56381411 миссенс вариант гена MAP3K15, rs9275373 в регуляторной области HLA-DQB1-MTCO3P1, rs148031082 в интроне гена PPP1R12A. Следует отметить, что rs36090025 гена TCF7L2 является уникальными для ГСД, а rs34872471, rs7903146 гена TCF7L2 ассоциированы как с СД2, так и с ГСД.

Анализ белок-белковых взаимодействий показал, что для белков, вовлеченных в патогенез ГСД, обогащение было статистически значимо ($p = 4.87e-06$), однако доказанных физических взаимодействий не было, и только упоминаются вместе в научных публикациях или коэкпрессируются (рис. 1, а). Для белков общих для ГСД и СД2 обогащение было также статистически значимо ($p < 1.0e-16$), при этом белки DMRTA1, CCND2 и CDKN2B имеют доказанное физическое взаимодействие, а остальные белки упоминаются вместе в научных публикациях или коэкпрессируются (см. рис. 1, б). Все белки, кроме DMRTA1, CCND2 и CDKN2B, вовлечены в развитие ГСД, нарушение углеводного обмена, наследственное нарушение обмена веществ, согласно базе данных DISEASES. Это свидетельствует о том, что данные белки находятся в одном молекулярном маршруте и требуется дальнейшее изучение их роли в развитии ГСД и СД2. Также они могут стать перспективными терапевтическими мишенями для разработки новых лекарственных препаратов.



а. **б.**

Рис. 1. Визуализация PPI для всех взаимодействия с обозначением белков их взаимодействия. Обозначения доступны в базе данных STRING; а – PPI проведенное для белков в области генов которых располагаются полиморфизмы, ассоциированные с СД2 и ГСД; б – PPI проведенное для белков в области генов которых располагаются полиморфизмы, ассоциированные с СД2 и ГСД

Выводы: Нами были выявлены общие для СД2 и ГСД и уникальные для ГСД ассоциации полиморфизмов. Большинство из 28 проанализированных полиморфизмов являются недавно выявленными и практическими неизученными. Расположение полиморфизмов в генах, белковые продукты которых совместно функционируют в общих молекулярных маршрутах создает перспективы как для дальнейшего изучения в рамках исследований случай-контроль, так и для разработки новой таргетной терапии.
Финансирование: Исследование выполнено в рамках государственного задания ФГБНУ «Научно-исследовательский институт патологии и патофизиологии» (№ ФГБНУ-2022-0011: «Выявление значимых биоиндикаторов различных нарушений функций организма»).

Identification of shared and distinct genetic variants associated with type 2 diabetes mellitus and gestational diabetes mellitus

Karpova N.S.*, Dmitrenko O.P., Nurbekov M.K.
Research Institute of General Pathology and Pathophysiology, Moscow, Russia
** nataliiakarpova.sp@gmail.com*

Key words: polymorphisms; GWAS; diabetes; gestational diabetes mellitus

Motivation and Aim: Gestational diabetes mellitus (GDM) is the most common extragenital pathology of pregnancy and represents a serious medical and social problem, increasing the incidence of adverse pregnancy outcomes for both the mother and the fetus [1]. A genetic factor has been identified in the etiopathogenesis of gestational diabetes mellitus. Poulsen et al. (2005) found that genetic factors explain more than 75 % of the dysfunction of pancreatic β-cells and glucose homeostasis and 53 % of the development of insulin resistance [2]. Various genetic analysis strategies,

such as genome-wide association searches (GWAS), are used to identify genetic variants (including single nucleotide polymorphisms (SNPs)) associated with GDM [3–5]. To date, about 40 genetic loci associated with GDM have been identified. Some of them are associated with the synthesis and secretion of insulin (INS, KCNJ11, ABCC8, TCF7L2, ND1), with insulin signal transmission (INSR, IGF2, IRS1) and regulation of carbohydrate and lipid metabolism (PPARG, PPARGC1A, ADRB3, GLUT1, ADIPOQ, FOXC2) [3]. Previous studies have shown that some genetic loci that predispose to the development of type 2 diabetes mellitus (T2DM) may also predispose to GDM [6]. Therefore, the aim of this study was to search for common and unique genetic variants associated with T2DM and GDM, followed by analysis of selected polymorphisms to discover key genes and pathways involved in the development of GDM and T2DM.

Methods and Algorithms: Information on polymorphisms associated with GDM and T2DM was obtained from the GWAS-catalog database (https://www.ebi.ac.uk/gwas/home) [3]. Linkage disequilibrium between polymorphisms was assessed using LDlink (Version 5.6.5, GRch38). The location of polymorphisms relative to the gene region was carried out in Ensembl Variant Effect Predictor (VEP) [7]. Protein-protein interaction (PPI) analysis was performed using STRING DB (Version 12.0, https://string-db.org) [8]. Genes that do not code for proteins were excluded from the analysis.

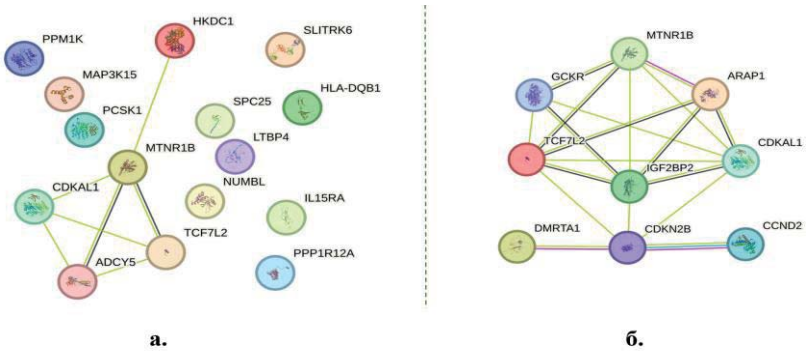
Results: Since 2019, 10 GWAS have been conducted and found 30 associations with GDM. Of these polymorphisms, 28 were identified in a single study, while the rs10830963 polymorphism in the MTNR1B gene was found to be associated with GDM in several independent studies (PMID: 36553520, 38182742, 35220425). According to LDlink (Version 5.6.5, GRch38), the rs10830962(C) allele correlates with the rs10830963(C) allele when analyzing all populations ($p < 0.0001$; D' = 0.9532; R2 = 0.2718), which indicates the benefit of more promising for further studies of the association of this genomic locus with GDM.

We analyzed the location of 28 polymorphisms relative to genetic regions using the Ensembl Variant Effect Predictor (VEP). The majority of polymorphisms are located in the intronic region of the gene (67.9 %). The remaining polymorphisms are located in the intergenic region (10.7 %), the gene region (missense variants; 7.1 %), the regulatory region (7.1 %), the non-coding variant of the transcript exon (3.6 %), the region is not defined (3.6 %). The location of polymorphisms in the regulatory zone may indicate their influence on gene expression, the formation of splicing variants of genes, as well as the functioning of regulatory RNAs. It was revealed that 13 polymorphisms were associated with both GDM and T2DM: It was revealed that 13 polymorphisms were associated with both GDM and T2DM: rs34872471, rs7903146 of the TCF7L2 gene intron, rs780094 (intron), rs1260326 (missense polymorphism), rs780093 (intron) GCKR gene, rs1470579, rs7633675 of the IGF2BP2 gene, rs9348441, rs7754840 intron of the CDKAL1 gene, rs76895963 in the non-coding exon of the CCND2-AS1 and CCND2 transcript, rs10811662 in the intergenic region of CDKN2B-AS1-DMRTA1, rs10830963 intron of the MTNR1B gene, rs77464186 of the intron of the ARAP1 gene.

A number of polymorphisms are associated only with GDM: rs6798189 of the ADCY5 gene, rs10220124 in the intergenic region of SLITRK6 – MOB1AP1, rs34499031 of the intron of the CDKAL1 gene, rs1402837 of the intron of the SPC25 gene, rs1820176 of the intron of the MIR583HG and PCSK1 gene, 589 intron of the PPM1K gene, rs36090025 intron of the TCF7L2 gene, rs7254268 in the regulatory region LTBP4-

NUMBL, rs10830962 in the intergenic region of SNRPGP16-MTNR1B, rs9663238 intron of the HKDC1 gene, rs2274034 intron of the IL15RA gene, rs56381411 missense variant of the MAP3K15 gene, rs9275373 in the regulatory region of HLA-DQB1-MTCO3P1, 8031082 in the intron of the PPP1R12A gene. It should be noted that rs36090025 of the TCF7L2 gene is unique to GDM, and rs34872471, rs7903146 of the TCF7L2 gene are associated with both T2DM and GDM.

Analysis of PPI showed that for proteins involved in the pathogenesis of GDM, the enrichment was statistically significant ($p = 4.87e-06$), but there were no proven physical interactions. and are only mentioned together in scientific publications or co-expressed (Fig. 1a). For proteins common to GDM and T2DM, the enrichment was also statistically significant ($p < 1.0e-16$), with the proteins DMRTA1, CCND2 and CDKN2B having a proven physical interaction, and the remaining proteins being mentioned together in scientific publications or being co-expressed (Fig. 1b). All proteins except DMRTA1, CCND2 and CDKN2B are involved in the development of GDM, carbohydrate metabolism disorders, and inherited metabolic disorders, according to the DISEASES database. This indicates that these proteins are located in the same molecular route and further study of their role in the development of GDM and T2DM is required. They can also become promising therapeutic targets for the development of new drugs.



а. **б.**

Fig. 1. Visualization of PPIs for all interactions with their interaction proteins labeled. Designations are available in the STRING database; a. PPI carried out for proteins in the gene region of which polymorphisms associated with T2DM and GDM are located; b. PPI carried out for proteins in the gene region of which polymorphisms associated with T2DM and GDM are located

Conclusion: We identified associations of polymorphisms common to T2DM and GDM and unique to GDM. Most of the 28 analyzed polymorphisms are recently identified and practically unstudied. The location of polymorphisms in genes whose protein products function together in common molecular pathways offers promise for both further study in case-control studies and for the development of new targeted therapies.

Funding: The study is supported within the framework of the state assignment of the Federal State Budgetary Institution “Research Institute of Pathology and Pathophysiology” (No. FGFU-2022-0011: “Identification of significant bioindicators of various disorders of body functions”).

Список литературы/References

1. Damm P. et al. Gestational diabetes mellitus and long-term consequences for mother and offspring: a view from Denmark. *Diabetologia*. 2016;59:1396-1399
2. Poulsen P. et al. Heritability of insulin secretion, peripheral and hepatic insulin action, and intracellular glucose partitioning in young and old Danish twins. *Diabetes*. 2005;54(1):275-283
3. Sollis E. et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023;51(D1):D977-D985
4. Dias S. et al. Molecular biomarkers for gestational diabetes mellitus. *Int J Mol Sci*. 2018;19(10):2926
5. Lu W. et al. Molecular biomarkers for gestational diabetes mellitus and postpartum diabetes. *Chin Med J*. 2022;135(16):1940-1951
6. Kawai V.K. et al. A genetic risk score that includes common type 2 diabetes risk variants is associated with gestational diabetes. *Clin Endocrinol*. 2017;87(2):149-155
7. McLaren W. et al. The ensembl variant effect predictor. *Genome Biol* 2016;17(1):122
8. Szklarczyk D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 2023;51(D1):D638-D646. doi 10.1093/nar/gkac1000

doi 10.18699/bgrs2024-9.2-10

Биоинформатический поиск низкомолекулярных лигандов гликированной формы человеческого сывороточного альбумина, способных модулировать его взаимодействие с β-амилоидным пептидом

Литус Е.[1]*, Шевелёва М.[1], Мачулин А.[2], Дерюшева Е.[1]

[1] *ФИЦ «Пущинский научный центр биологических исследований» РАН, Институт биологического приборостроения, Пущино, Московская область, Россия*

[2] *ФИЦ «Пущинский научный центр биологических исследований» РАН, Институт биохимии и физиологии микроорганизмов им. Г.К. Скрябина, Пущино, Московская область, Россия*

** ealitus@gmail.com*

Ключевые слова: человеческий сывороточный альбумин; гликированная форма; β-амилоидный пептид; низкомолекулярные лиганды; модулирование взаимодействий

Мотивация и цель: Болезнь Альцгеймера (БА) представляет собой серьезную проблему в обществе, являясь одной из наиболее распространенных форм деменции. Смертность от БА стремительно растет: по данным ВОЗ, она входит в десятку основных причин смертности в мире. Ключевым механизмом развития БА является накопление β-амилоидного пептида (Аβ) в тканях головного мозга, что приводит к образованию амилоидных бляшек. Этот процесс вызывает нарушение нейронной функции и приводит к постепенной деградации когнитивных способностей [1]. Исследования показывают, что увеличение скорости выведения Аβ может использоваться для замедления прогрессирования БА. Один из перспективных методов в данном направлении – воздействие на человеческий сывороточный альбумин (ЧСА), который играет биологически значимую роль в транспортировке Аβ [2]. ЧСА связывает около 90 % Аβ в сыворотке крови, а в цереброспинальной жидкости от 40 до 94 %. ЧСА подавляет агрегацию Аβ и снижает риск развития БА и ее прогрессию. По этой причине ЧСА рассматривается как терапевтическая мишень для лечения БА. Одним из разрабатываемых подходов в данном направлении является увеличение сродства ЧСА к Аβ путем изменения концентрации низкомолекулярных лигандов ЧСА в крови. Наши исследования показывают возможность увеличения сродства ЧСА к Аβ с помощью таких низкомолекулярных лигандов ЧСА как линолевая кислота, серотонин, ибупрофен и преднизон. Кроме того, некоторые из лигандов ЧСА (рисперидон, мефенаминовая кислота) стоит рассматривать как потенциально опасные. Наши данные показывают, что они способны уменьшать константу связывания мономерной формы Аβ с ЧСА. Это может привести к увеличению «свободной» фракции Аβ, что в свою очередь способствует образованию амилоидных фибрилл [3]. Отметим, что полученные нами данные *in vitro* согласуются с результатами клинических и эпидемиологических исследований. Для дальнейшего развития данного направления необходимо проведение экспериментов в условиях, более полно соответствующих условиям *in vivo*. Так, например, не изучено влияния модификаций самого ЧСА на его функциональную активность по отношению к Аβ. Известно, что уровень гликированной фракции

ЧСА повышен у пациентов с диагнозом БА [4]. Уровень гликированного ЧСА в сыворотке пациентов с сахарным диабетом, ассоциированным с БА, может достигать 30 % от общей фракции ЧСА [5]. В данной работе проведен поиск низкомолекулярных лигандов гликированной формы ЧСА, способных модулировать его взаимодействие с Аβ.

Методы и алгоритмы: Для формирования панели низкомолекулярных веществ – лигандов гликированной фракции ЧСА, потенциально способных модулировать образование комплекса ЧСА-Аβ, в качестве основного источника данных была использована база данных DrugBank (https://go.drugbank.com). Алгоритм выбора веществ подробно описан в [3]. Отметим, что в качестве критериев отбора были использованы физико-химические характеристики веществ, а также их строгая ассоциация с БА. Конечная панель веществ содержит 100 наименований. Для оценки специфичности сайтов связывания исследуемых лигандов были смоделированы трехмерные комплексы низкомолекулярных лигандов с ЧСА. Трехмерная структура ЧСА была получена из Protein Data Bank (https://www.rcsb.org, PDB ID: 1ao6, chain A). Для подготовки ЧСА к моделированию использовалось программное обеспечение AutoDockTools (ADT) (http://autodock.scripps.edu/resources/adt/index_html). Трехмерные (3D) структуры панели лигандов были получены в формате sdf с сервера PubChem (https://www.ncbi.nlm.nih.gov/pccompound) и преобразованы в файлы pdb с помощью PyMOL v.1.6.9.0. (https://pymol.org/2/). Программа Auto Dock Vina (http://vina.scripps.edu/index.html) использовалась для молекулярного докинга выбранных лигандов с ЧСА. Комплексы белок-лиганд были визуализированы с помощью PyMOL v. 1.6.9.0. (https://pymol.org/2/). Характеристики взаимодействий в комплексах белок-лиганд были получены с помощью специализированного сервиса PLIP (https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index). Для каждого комплекса лиганд-ЧСА были взяты три модели, соответствующие комплексам с наименьшей энергией. Число рассматриваемых далее комплексов было выбрано оптимальным, исходя из анализа литературных данных и доступных трехмерных структур комплексов ЧСА-низкомолекулярный лиганд: у одного и тоже лиганда число сайтов связывания варьируется от одного до трех [6]. Для каждого вещества панели (по трем моделям) был рассчитан процент содержания аминокислот, входящий в сайт связывания. Реализация обработки полученных данных была осуществлена на свободно распространяемом высокоуровневом языке программирования Python 3.6, а также языке статистической обработки данных R.

Результаты: Известно, что N-конец и 59 остатков лизина могут выступать в качестве потенциальных мест для образования гликированной формы ЧСА [7, 8]. Также, ЧСА содержит 24 остатка аргинина, которые, наряду с лизинами и N-концом, потенциально могут быть задействованы в формировании гликированной формы ЧСА [7]. Анализ наличия в сайтах связывания веществ нашей панели низкомолекулярных лигандов (по трем моделям) остатков лизина и аргинина позволил выявить вещества с наименьшим содержанием этих остатков в сайтах связывания с ЧСА: тестостерон (Drug Bank ID: DB00624, 7.4 % Lys, 3.7 % Arg), флурбипрофен (Drug Bank ID: DB00712, 11.1 % Lys, 7.4 % Arg), абиратерон (Drug Bank ID: DB05812, 23.1% Lys, 0 % Arg), эналаприл (Drug Bank ID: DB00584, 14.8% Lys, 11.1% Arg), фторурацил (Drug Bank ID: DB00544, 11.1% Lys, 14.8% Arg), салициловая кислота (Drug Bank ID: DB00936, 14.8% Lys, 14.8 % Arg), эстрадиол (Drug Bank ID: DB00783, 18.5 % Lys, 11.1 % Arg), метопролол (Drug Bank ID:

DB00264, 7.4 % Lys, 29.6 % Arg). Можно предположить, что эти лиганды в большей степени смогут модулировать взаимодействие гликированной формы ЧСА с Aβ вследствие уменьшения конкуренции за связывание ЧСА с глюкозой в сравнении с теми лигандами, центр связывания которых затронут гликированием. Отметим, что выбранные вещества по полученным моделям связаны с разными участками ЧСА и, вероятно, могут по-разному влиять на взаимодействие модифицированной формы ЧСА с Aβ.
Выводы: Таким образом, в данной работе были выбраны 8 лекарственных препаратов, потенциально способных модулировать взаимодействие гликированной формы ЧСА с Aβ. Полученные данные могут использоваться для рационального перепрофилирования лекарственных препаратов при разработке персонализированного подхода к лечению и профилактики БА у пациентов с сопутствующим сахарным диабетом.
Финансирование: Исследование поддержано грантом Российского научного фонда (№ 20-74-10072, https://rscf.ru/project/20-74-10072/).

Bioinformatic search for low-molecular-weight ligands ligands of the glycated form of human serum albumin capable of modulating its interaction with amyloid β peptide

Litus E.[1]*, Shevelyova M.[1], Machulin A.[2], Deryusheva E.[1]
[1] *Institute for Biological Instrumentation, Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences, Pushchino, Russia*
[2] *Skryabin Institute of Biochemistry and Physiology of Microorganisms, Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences, Pushchino, Russia*
* *ealitus@gmail.com*

Key words: human serum albumin; glycated form; amyloid β peptide; low-molecular-weight ligands; modulation of interactions

Motivation and Aim: Alzgeimer desease (AD) is one of the most prevalent dementia forms that pose a serious society problem. According to WHO data mortality from AD increases rapidly and now it is one of the ten leading causes of death in the world. The key mechanism of AD development is amyloid β peptide (Aβ) accumulation in brain tissue leading to amyloid plague formation. The process causes neuronal dysfunction and leads to gradual degradation of cognitive abilities [1]. It was demonstrated earlier that increasing the rate of Aβ clearance may be used to slowing down AD progression. One of the perspective methods is the influence on human serum albumin (HSA) which plays biologically important role in the Aβ transport [2]. HSA binds about 90 % of Aβ in serum and decreases AD risk and its development. Due to this HSA is considered as therapeutic target during AD treatment. One of the approaches being developed in this direction is increase of HSA-Aβ affinity by changing of low-molecular weight HSA ligands concentration in plasma. Our recent studies demonstrate possibility of the affinity increase using such low-molecular weight HSA ligands as linoleic acid, serotonin, ibuprofen and prednisone. Moreover, some of the HSA ligands (risperidone, mefenamic acid) should be considered as potentially dangerous. Our data show that they

are capable of reducing the binding constant of the monomeric Aβ to HSA. This can lead to an increase of the "free" Aβ fraction, which in turn promotes the formation of amyloid fibrils [3]. This *in vitro* data agree with the results of clinical and epidemiological studies. For further development, the experiment conditions should more fully correspond to *in vivo* conditions. For example, the effect of HSA modifications on its functional activity towards Aβ has not been studied. It is known that the level of the glycated HSA fraction is increased in AD patients [4]. The level of glycated HSA in the serum of patients with diabetes mellitus associated with AD can reach 30% of the total HSA fraction [5]. In this work, we searched for low-molecular weight ligands of the glycated HSA that can modulate its interaction with Aβ.

Methods and Algorithms: To form a panel of low-molecular weight ligands of the glycated HSA potentially capable of modulating the HSA-Aβ complex formation, the DrugBank database (https://go.drugbank.com) was used as the main data source. The selecting algorithm is described in detail in [3]. The physicochemical characteristics of the substances, as well as their strong association with AD, were used as selection criteria. The final panel of substances contains 100 items. In order to evaluate the specificity of the binding sites of the studied ligands, three-dimensional complexes of the ligands with HSA were modeled. The three-dimensional structure of HSA was obtained from the Protein Data Bank (https://www.rcsb.org, PDB ID: 1ao6, chain A). AutoDockTools (ADT) software (http://autodock.scripps.edu/resources/adt/index_html) was used to prepare the HSA for modeling. Three-dimensional (3D) structures of the ligands were obtained in sdf format from the PubChem server (https://www.ncbi.nlm.nih.gov/pccompound) and converted to pdb files using PyMOL v.1.6.9.0. (https://pymol.org/2/). The Auto Dock Vina program (http://vina.scripps.edu/index.html) was used for molecular docking of the selected ligands with HSA. Protein-ligand complexes were visualized using PyMOL v. 1.6.9.0. (https://pymol.org/2/). Characteristics of interactions in protein-ligand complexes were obtained using the specialized PLIP service (https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index). For each ligand-HSA complex, three models with the lowest energy were taken. The number of complexes considered below was chosen as optimal based on an analysis of literature data and available three-dimensional structures of HSA-low molecular weight ligand complexes: for each ligand, the number of binding sites varies from one to three [6]. For each substance in the panel (according to three models), the percentage of amino acid content included in the binding site was calculated. The implementation of processing the obtained data was carried out in the freely distributed high-level programming language Python 3.6, as well as the statistical data processing R language.

Results: N-terminus and 59 lysine residues of HSA are known to be potential sites for glycation [7, 8]. Moreover, HSA contains 24 arginine residues which could also be involved in the formation of the glycated form [7]. Analysis of the presence of lysine and arginine residues in the binding sites of the low-molecular weight ligands involved in the panel (according to three models) allowed us to identify substances with the lowest content of these residues in the HSA binding sites: testosterone (Drug Bank ID: DB00624, 7.4 % Lys, 3.7 % Arg), flurbiprofen (Drug Bank ID: DB00712, 11.1 % Lys, 7,4 % Arg), abiraterone (Drug Bank ID: DB05812, 23.1 % Lys, 0 % Arg), enalapril (Drug Bank ID: DB00584, 14.8 % Lys, 11.1 % Arg), fluorouracil (Drug Bank ID: DB00544, 11.1 % Lys, 14.8 % Arg), salicylic acid (Drug Bank ID: DB00936, 14.8 % Lys, 14.8 % Arg), estradiol (Drug Bank ID: DB00783, 18.5 % Lys, 11.1 % Arg),

metoprolol (Drug Bank ID: DB00264, 7.4 % Lys, 29.6 % Arg). It is worth to assume that according to the obtained models the chosen ligands bind to different regions of HSA molecule and, consequently, may have different effect on the interaction of modified HSA and Aβ.

Conclusion: 8 drugs potentially capable of modulating interaction between glycated HSA and Aβ were chosen in this work. The obtained data can be used for repurposing of drugs in the development of personalized approach to the treatment and prevention AD in patients with concomitant diabetes mellitus.

Funding: The study is supported by the Russian Science Foundation (No. 20-74-10072, https://rscf.ru/project/20-74-10072/).

Список литературы/References

1. Boada M., Kiprov D., Anaya F. et al. Feasibility, safety, and tolerability of two modalities of plasma exchange with albumin replacement to treat elderly patients with Alzheimer's disease in the AMBAR study. *J Clin Apher*. 2023;38(1):45-54. doi 10.1002/jca.22026
2. Shevelyova M.P., Deryusheva E.I., Nemashkalova E.L. et al. Role of human Serum albumin in the prevention and treatment of Alzheimer's Disease. *Biol Bull Rev*. 2024;14(1):29-42. doi 10.1134/S2079086424010109
3. Deryusheva E.I., Shevelyova M.P., Rastrygina V.A. et al. In search for Low-Molecular-weight Ligands of human Serum albumin That affect Its affinity for monomeric Amyloid β Peptide. *Int J Mol Sci*. 2024;25(9):4975. doi 10.3390/ijms25094975
4. Costa M., Mestre A., Horrillo R. et al. Cross-sectional characterization of albumin glycation state in cerebrospinal fluid and plasma from Alzheimer's disease patients. *J Prev Alzheimer's Dis*. 2019;6(2):139-143. doi 10.14283/jpad.2018.48
5. Wu W.C., Ma W.Y., Wei J.N. et al. Serum Glycated Albumin to Guide the Diagnosis of Diabetes Mellitus. *PLoS One*. 2016;11(1):e0146780. doi 10.1371/journal.pone.0146780
6. Ghuman J., Zunszain P.A., Petitpas I. et al. Structural basis of the drug-binding specificity of human serum albumin. *J Mol Biol*. 2005;353:38-52. doi 10.1016/j.jmb.2005.07.075
7. Shaklai N., Garlick R.L., Bunn H. Nonenzymatic glycosylation of human serum albumin alters its conformation and function. *J Biol Chem*. 1984;259(6):3812-17. doi 10.1016/S0021-9258(17)43168-1
8. Anguizola J., Matsuda R., Barnaby O.S. et al. Review: glycation of human serum albumin. *Clinica Chimica Acta*. 2013;425:64-76. doi 10.1016/j.cca.2013.07.013

doi 10.18699/bgrs2024-9.2-11

Новые конъюгаты фосфорилгуанидиновых олигонуклеотидов с остатками бис-имидазол-содержащих пептидоподобных соединений

Павлова А.*, Дюдеева Е., Королева Л., Григорьева Е., Дмитриенко Е., Коваль О., Пышный Д., Пышная И.

Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** pavlova@niboch.nsc.ru*

Ключевые слова: антисмысловые олигонуклеотиды; фосфорилгуанидины; CuAAC; расщепление РНК

Мотивация и цель: Присоединение низкомолекулярных реакционно-способных соединений, расщепляющих фосфодиэфирные связи в РНК, к антисмысловым олигонуклеотидам является актуальным направлением в биомедицине для разработки новых средств на их основе и/или повышения эффективности их действия [1]. Отдельную задачу представляет синтез конъюгатов модифицированных олигонуклеотидов с РНК-расщепляющими группировками, поскольку это может расширить возможности их применения. Целью данной работы было получение конъюгатов олигодезоксирибонуклеотидов, в том числе содержащих фосфорилгуанидиновые (ФГ) модификации, с остатками бис-имидазол-содержащих пептидоподобных соединений и исследование их активности *in vitro*.

Методы и алгоритмы: Для синтеза конъюгатов использовали реакцию азид-алкинового циклоприсоединения, катализируемую ионами одновалентной меди, протекающую с образованием 1,4-дизамещенных 1,2,3-триазолов (CuAAC). Реакцию проводили с участием олигонуклеотида, закреплённого на полимерном носителе CPG и содержащего на 5'-конце терминальную алкинильную группировку. Предварительное введение в олигонуклеотид алкинильной группы проводили в процессе стандартного твердофазного фосфитамидного синтеза с использованием ранее полученных соответсвующих амидофосфитов [2]. После выделения синтезированных конъюгатов методом электрофореза проводили исследование их реакционной способности к гидролизу РНК-мишени в условиях, близких к физиологическим. Способность конъюгатов к подавлению экспрессии мРНК проводили с использованием клеточной культуры HEK293-FT, экспрессирующей EGFP, методом цитофлуориметрии по изменению уровня флуоресценции белка. Для трансфекции олигонуклеотидов и их конъюгатов использовали Lipofectamine® 3000.

Результаты: Синтезирован ряд конъюгатов олигодезоксирибонуклеотидов, в том числе содержащих ФГ модификации, с остатками бис-имидазол-содержащих пептидоподобных соединений, отличающихся длиной линкеров в своей структуре. Исследование реакционной способности показало, что в случае использования конъюгатов с десятью ФГ модификациями короткие модельные РНК гидролизуются *in vitro* в среднем на 8–20 % лучше (от ~16 до ~50 % за 3 ч), чем в случае конъюгатов без модификаций (от ~8 до ~30 %). Наибольшая степень

расщепления мишени (~50 %) наблюдалась для ФГ-содержащего конъюгата с наиболее коротким линкером между остатками имидазола в присоединенной группировке. Исследование подавления экспрессии мРНК EGFP показало, что в сравнении с олигонуклеотидом без реакционно-способной группировки конъюгаты демонстрировали схожую эффективность по снижению уровня флуоресценции экспрессируемого клетками белка. Степень подавления уровня EGFP различалась в зависимости от используемой последовательности антисмыслового олигонуклеотида или конъюгата.
Выводы: Таким образом, с использованием ранее полученных амидофосфитов предложен твердофазный вариант применения реакции CuAAC для синтеза конъюгатов олигодезоксирибонуклеотидов, в том числе содержащих ФГ модификации, с остатками бис-имидазол-содержащих пептидоподобных соединений, и проведена валидация методики для ряда нуклеотидных последовательностей. С использованием коротких РНК-мишеней in vitro показано, что введение в олигонуклеотид ФГ модификаций позволяет повысить степень расщепления РНК данными конъюгатами. С использованием клеточной культуры показано, что присоединение бис-имидазол-содержащих пептидомиметиков не оказывает негативного влияния на эффективность подавления уровня экспрессии EGFP.
Финансирование: Исследование проводилось в рамках государственного задания ИХБФМ СО РАН № 121031300042-1.

Novel conjugates of phosphorylguanidine oligonucleotides with bisimidazol-containing peptidomimetics

Pavlova A.*, Dyudeeva E., Koroleva L., Grigoryeva E., Dmitrienko E., Koval O., Pyshnyi D., Pyshnaya I.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** pavlova@niboch.nsc.ru*

Key words: antisense oligonucleotides; phosphorylguanidines; CuAAC; RNA cleavage

Motivation and Aim: The attachment of low-molecular-weight bioreactive compounds that cleave phosphodiester bonds in RNA to antisense oligonucleotides is a relevant area in biomedicine for the development of new agents based on them and/or increasing their efficacy [1]. Synthesis of conjugates of modified oligonucleotides with RNA-cleaving groups represents a special task, as it can expand the possibilities of their application. The aim of this work was to synthesise conjugates of oligodeoxyribonucleotides, including those containing phosphorylguanidine (PG) modifications, with residues of bisimidazole-containing peptidomimetics and to study their activity in vitro.
Methods and Algorithms: For the synthesis of conjugates, the azide-alkyne cycloaddition catalysed by monovalent copper ions was used to form 1,4-disubstituted 1,2,3-triazoles (CuAAC). The reaction was carried out with an oligonucleotide that had a terminal alkynyl group at the 5' end and was attached to a CPG solid support. The introduction of the alkynyl group into the oligonucleotide was carried out by standard phosphoramidite solid-phase synthesis using previously obtained corresponding phosphoramidites [2]. After isolation of the synthesized conjugates, their reactivity towards the hydrolysis of

the target RNA was studied by electrophoresis under conditions close to physiological. The ability of the conjugates to inhibit mRNA expression was tested in HEK293-FT cell cultures expressing EGFP using flow cytometry to determine the change in protein fluorescence levels. The oligonucleotides and their conjugates were transfected using the Lipofectamine® 3000.

Results: A series of conjugates of oligodeoxyribonucleotides, including those containing PG modifications, with residues of bisimidazole-containing peptidomimetics differing in the length of linkers in their structure were synthesized. Reactivity studies showed that conjugates with ten PG modifications hydrolyzed short model RNAs in vitro on average 8–20% better (from ~16 to ~50 % in 3 h) than conjugates without modifications (from ~8 to ~30 %). The highest degree of target cleavage (~50 %) was observed for the PG-containing conjugate with the shortest linker between imidazole residues in the attached moiety. EGFP mRNA silencing studies showed that the conjugates were similarly effective in reducing the level of fluorescence of the protein expressed by the cells compared to an oligonucleotide without a reactive group. The degree of EGFP silencing varied depending on the antisense oligonucleotide or conjugate sequence used.

Conclusion: Thus, using previously obtained phosphoramidites, a solid-phase version of the CuAAC reaction was proposed for the synthesis of conjugates of oligodeoxyribonucleotides, including those containing PG modifications, with residues of bisimidazole-containing peptidomimetics, and the method was validated for a number of nucleotide sequences. Using short RNA targets in vitro, it was shown that the introduction of PG modifications to the oligonucleotide can increase the degree of RNA cleavage by these conjugates. Using cell culture, it has been shown that the attachment of bisimidazole-containing peptidomimetics to the antisense oligonucleotides does not adversely affect the efficiency of silencing EGFP expression levels.

Funding: This work was financially supported by the Russian State-funded project for the Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences (ICBFM, SB RAS) (grant number 121031300042-1).

Список литературы/References

1. Staroseletz Y., Gaponova S., Patutina O., Bichenkova E., Amirloo B., Heyman T., Chiglintseva D., Zenkova M. Site-Selective Artificial Ribonucleases: Renaissance of Oligonucleotide Conjugates for Irreversible Cleavage of RNA Sequences. *Molecules.* 2021;26(6):1732. doi 10.3390/molecules26061732
2. Pavlova A.S., Ogurtsova P.A., Koroleva L.S., Serpokrylova I.Y., Lomzov A.A., Pyshnaya I.A., Silnikov V.N., Pyshnyi D.V. Novel Bisimidazole-Containing Peptidomimetic Molecules for Metal-Independent RNA Cleavage: Synthesis and Solid-Phase Screening Method. *Rus. J. Bioorg. Chem.* 2019;45(6):813-824. doi 10.1134/S1068162019060311

doi 10.18699/bgrs2024-9.2-12

Оценка влияния добавок жирных кислот омега-3 (SolowaysTM) на липидный профиль у взрослых с полиморфизмом PPARG: рандомизированное двойное слепое плацебо-контролируемое исследование

Покушалов Е.[1, 2], Пономаренко А.[1]*, Байрамова С.[1], Гарсия К.[2], Пак И.[1], Шрайнер Е.[1], Джонсон М.[2], Миллер Р.[2]

[1] *Центр новых медицинских технологий, Новосибирск, Россия*

[2] *Научно-исследовательская лаборатория Triangel Scientific, Сан-Франциско, Калифорния, США*

* *dayshadoff@gmail.com*

Ключевые слова: полиморфизмы PPARG; жирные кислоты омега-3; ЛПНП; триглицериды; сердечно-сосудистые заболевания; персонализированная медицина

Мотивация и цель: Сердечно-сосудистые заболевания (ССЗ) остаются основной причиной заболеваемости и смертности во всем мире, что подчеркивает необходимость эффективных стратегий профилактики и лечения [1]. Дислипидемия, характеризующаяся повышенным уровнем холестерина ЛПНП и триглицеридов при снижении уровня холестерина ЛПВП, играет ключевую роль в патогенезе ССЗ [2].

Управление уровнями холестерина важно для снижения рисков ССЗ, особенно у пациентов с низким и умеренным сердечно-сосудистым риском [2]. Для этих пациентов часто рекомендуются нефарамакологические стратегии, включая пищевые добавки [2]. Мутации в гене PPARG, ключевом регуляторе хранения жирных кислот и метаболизма глюкозы, связаны с повышенным сердечно-сосудистым риском [3].

Омега-3 полиненасыщенные жирные кислоты, в основном получаемые из морских масел, демонстрируют эффективность в улучшении сердечной функции и снижении воспалительных процессов [4, 5]. Эти жирные кислоты могут модулировать экспрессию PPARG, снижая уровни ЛПНП и уменьшая сердечно-сосудистый риск [6].

Это исследование было направлено на оценку эффективности добавок рыбьего жира в улучшении сердечно-сосудистых показателей у взрослых с подтвержденными полиморфизмами гена PPARG [7, 8].

Методы и алгоритмы: это рандомизированное двойное слепое параллельное клиническое исследование сравнивало лечение омега-3 жирными кислотами с плацебо. Все пациенты предоставили письменное информированное согласие. Исследование зарегистрировано на ClinicalTrials.gov (NCT06154408). Критериями включения были: возраст 40–75 лет, уровень ЛПНП 70–190 мг/дл.

Пациенты были рандомизированы на группы омега-3 жирных кислот (n = 51) и плацебо (n = 51). Все участники принимали 2 капсулы в день. Исследование длилось 90 дней, измерялись липидный профиль, метаболическая панель и С-реактивный белок высокой чувствительности на 0 и 90-й день.

Первичной конечной точкой было изменение уровня ЛПНП у пациентов с полиморфизмом гена PPARG. Вторичные конечные точки включали изменения

уровней С-реактивного белка, холестерина ЛПВП, общего холестерина и триглицеридов между подгруппами.

Результаты: Всего 102 пациента были рандомизированы, 99 успешно завершили исследование. Соблюдение режима было высоким, только четыре участника приняли менее 70 % назначенных препаратов. Частота аллеля для rs10865710 (G), rs7649970 (T), rs1801282 (G) и rs3856806 (T) составила 0.201, 0.088, 0.108 и 0.127 соответственно, и все SNP соответствовали равновесию Харди–Вайнберга (p > 0,05). Основные характеристики пациентов не различались между группами омега-3 и плацебо. У пациентов с полиморфизмом PPARG были более высокие уровни ЛПНП на исходном уровне (135.3 ± 23.2 против 123.7 ± 19.3 мг/дл, p = 0.04 в группе омега-3 и 139.2 ± 27.1 против 119.9 ± 17.8 мг/дл, p = 0.02 в группе плацебо). Также различались уровни ЛПВП и общего холестерина.

Основная конечная точка: уменьшение уровня ЛПНП на 15.4 % в группе омега-3 с полиморфизмом PPARG против 2.6 % в группе плацебо (разница 12,8 %, p < 0.01) [1]. У пациентов без полиморфизма PPARG снижение уровня ЛПНП было менее выраженным (3.7 % в группе омега-3 против 2.9 % в группе плацебо, p = 0,28) [2].

Вторичные конечные точки: уменьшение уровня триглицеридов на 21.3 % в группе омега-3 с полиморфизмом PPARG против 1.9 % в группе плацебо (p < 0.01). Изменения общего холестерина, ЛПВП и hsCRP не были значительными (p > 0.05 для всех).

Выводы: Исследование доказывает важную роль полиморфизмов PPARG в липидном обмене при приеме омега-3 жирных кислот. Результаты рандомизированного двойного слепого исследования показывают, что пациенты с полиморфизмами PPARG и повышенным холестерином получают значительные преимущества от добавления омега-3 в рацион. Эти преимущества включают снижение уровня ЛПНП, общего холестерина и триглицеридов.

Выводы подчеркивают потенциал персонализированного питания, предлагая использовать генетическое профилирование для оптимизации липидоснижающих вмешательств. Пациенты с полиморфизмами PPARG показали лучшие улучшения липидного профиля, что свидетельствует о необходимости учитывать генетическую предрасположенность при назначении омега-3 добавок. Будущие исследования должны изучить долгосрочные эффекты омега-3 добавок и молекулярные механизмы взаимодействия вариантов PPARG и липидного обмена.

Финансирование: Авторы заявляют, что это исследование получило финансирование от SolowaysTM. Спонсор не участвовал в разработке исследования, сборе, анализе, интерпретации данных, написании или решении предоставлять результатов исследования для публикации.

Evaluation of the effect of omega-3 fatty acid supplementation (SolowaysTM) on lipid profile in adults with PPARG polymorphism: a randomized, double-blind, placebo-controlled trial

Pokushalov E.[1, 2], Ponomarenko A.[1]*, Bayramova S.[1], Garcia K.[2], Pak I.[1], Shreiner E.[1], Johnson M.[2], Miller R.[2]

[1] *Center for New Medical Technologies, Novosibirsk, Russia*

[2] *Triangel Scientific Research Laboratory, San Francisco, CA, USA*

* *dayshadoff@gmail.com*

Key words: PPARG polymorphisms; omega-3 fatty acids; LDL-C; triglycerides; cardiovascular health; personalized medicine

Motivation and Aim: Cardiovascular diseases (CVDs) remain a leading cause of morbidity and mortality worldwide, underscoring the need for effective preventive and therapeutic strategies [1]. Dyslipidemia, characterized by elevated levels of LDL cholesterol and triglycerides with reduced HDL cholesterol levels, plays a key role in the pathogenesis of CVDs [2].

Managing cholesterol levels is crucial for reducing CVD risks, especially in patients with low to moderate cardiovascular risk [2]. Non-pharmacological strategies, including dietary supplements, are often recommended for these patients [2]. Mutations in the PPARG gene, a key regulator of fatty acid storage and glucose metabolism, are associated with increased cardiovascular risk [3].

Omega-3 polyunsaturated fatty acids, primarily derived from marine oils, have shown effectiveness in improving cardiac function and reducing inflammatory processes [4, 5]. These fatty acids can modulate the expression of PPARG, lowering LDL levels and reducing cardiovascular risk [6].

This study aimed to evaluate the effectiveness of fish oil supplements in improving cardiovascular markers in adults with confirmed PPARG gene polymorphisms [7, 8].

Methods and Algorithms: This randomized double-blind parallel clinical trial compared the treatment of omega-3 fatty acids with a placebo. All patients provided written informed consent. The study is registered on ClinicalTrials.gov (NCT06154408). Inclusion criteria were: age 40-75 years, LDL level 70–190 mg/dL.

Patients were randomized into omega-3 fatty acids ($n = 51$) and placebo ($n = 51$) groups. All participants took 2 capsules per day. The study lasted 90 days, with lipid profile, metabolic panel, and high-sensitivity C-reactive protein measured on days 0 and 90.

The primary endpoint was the change in LDL levels in patients with the PPARG gene polymorphism. Secondary endpoints included changes in C-reactive protein levels, HDL cholesterol, total cholesterol, and triglycerides between subgroups.

Results: A total of 102 patients were randomized, with 99 successfully completing the study. Adherence to the regimen was high, with only four participants taking less than 70 % of their assigned medications. The allele frequency for rs10865710 (G), rs7649970 (T), rs1801282 (G), and rs3856806 (T) was 0.201, 0.088, 0.108, and 0.127 respectively, and all SNPs were in Hardy-Weinberg equilibrium ($p > 0.05$).

The baseline characteristics of patients did not differ between the omega-3 and placebo groups. Patients with the PPARG polymorphism had higher baseline LDL levels (135.3 ± 23.2 vs. 123.7 ± 19.3 mg/dL, $p = 0.04$ in the omega-3 group and 139.2 ± 27.1 vs. 119.9 ± 17.8 mg/dL, $p = 0.02$ in the placebo group). HDL and total cholesterol levels also differed.

Primary endpoint: LDL levels decreased by 15.4 % in the omega-3 group with PPARG polymorphism compared to 2.6 % in the placebo group (difference of 12.8 %, $p < 0.01$) [1]. In patients without the PPARG polymorphism, LDL reduction was less pronounced (3.7 % in the omega-3 group vs. 2.9 % in the placebo group, $p = 0.28$) [2].

Secondary endpoints: triglyceride levels decreased by 21.3 % in the omega-3 group with PPARG polymorphism compared to 1.9 % in the placebo group ($p < 0.01$). Changes in total cholesterol, HDL, and hsCRP were not significant ($p > 0.05$ for all).

Conclusion: The study demonstrates the important role of PPARG polymorphisms in lipid metabolism when taking omega-3 fatty acids. Results from the randomized double-blind trial show that patients with PPARG polymorphisms and elevated cholesterol significantly benefit from adding omega-3 to their diet. These benefits include reductions in LDL, total cholesterol, and triglycerides.

The findings highlight the potential of personalized nutrition, suggesting the use of genetic profiling to optimize lipid-lowering interventions. Patients with PPARG polymorphisms showed better improvements in lipid profiles, indicating the need to consider genetic predispositions when prescribing omega-3 supplements. Future research should explore the long-term effects of omega-3 supplements and the molecular mechanisms of the interaction between PPARG variants and lipid metabolism.

Funding: The authors declare that this research received funding from SolowaysTM. The funder had no role in study design, data collection, analysis, interpretation, writing, or the decision to submit the study for publication.

Список литературы/References

1. Roth G.A. et al. Global burden of cardiovascular diseases and risk factors, 1990-2019, update from the GBD 2019 study. *J Am Coll Cardiol*. 2020;76:2982-3021
2. Song Y. et al. PPARγ gene polymorphisms, metabolic disorders, and coronary artery disease. *Front Cardiovasc Med*. 2022;9:808929. doi 10.3389/fcvm.2022.808929
3. Sherratt S.C.R. et al. Role of Omega-3 fatty acids in cardiovascular disease: the debate continues. *Curr Atheroscler Rep*. 2023;25(1):1-17. doi 10.1007/s11883-022-01075-x
4. Von Schacky C., Harris W.S. Cardiovascular benefits of omega-3 fatty acids. *Cardiovasc Res*. 2007;73:310-315
5. Bhat S. et al. Omega-3 Fatty acids in cardiovascular disease and diabetes: a review of recent evidence. *Curr Cardiol Rep*. 2023;25:51-65
6. Ghasemi Darestani N. et al. Association of polyunsaturated fatty acid intake on inflammatory gene expression and multiple sclerosis: a systematic review and meta-analysis. *Nutrients*. 2022;14(21):4627. doi 10.3390/nu14214627
7. Chaddha A. et al. Omega-3 fatty acids and heart health. *Circulation*. 2015;132:e350-e352
8. Bowen K.J. et al. Omega-3 fatty acids and cardiovascular disease: are there benefits? *Curr Treat Options Cardiovasc Med*. 2016;18(11):69. doi 10.1007/s11936-016-0487-1

doi 10.18699/bgrs2024-9.2-13

Оценка влияния репродуктивной патологии на количество копий TREC/KREC в «сухой капле» крови у недоношенных новорожденных

Полякова Е.[1]*, Берестень С.[2], Белевцев М.[1]

[1] *Республиканский научно-практический центр детской онкологии, гематологии и иммунологии, Минск, Беларусь*

[2] *Республиканский научно-практический центр «Мать и дитя», Минск, Беларусь*

** polyakovakat86@gmail.com*

Ключевые слова: репродуктивная патология; недоношенные новорожденные; ПЦР в «реальном времени»; TREC/KREC

Мотивация и цель: Иммунологический статус новорожденного ребенка формируется в условиях иммунологических взаимоотношений матери и плода. Различные виды репродуктивной патологии матери нарушают эти взаимоотношения, что может оказать влияние не только на внутриутробное развитие плода, но и на его иммунную систему. Инфекционно-воспалительные осложнения беременности повышают риск преждевременных родов. В результате чего рождаются недоношенные дети, у которых иммунная система отличается от доношенных новорожденных своей функциональной незрелостью. Механизмы изменений врожденной иммунной системы и влияние незрелости на общий риск развития иммунологических нарушений, связанный с высоким риском раннего развития инфекционной патологии в неонатальном периоде, до конца не изучены, но наибольшим риском считают преждевременные роды.

Методы и алгоритмы: Мы исследовали уровни маркеров неогенеза Т-и В-лимфоцитов (TREC и KREC) у 100 недоношенных детей со сроком беременности 36–37 недель посредством ПЦР в «реальном времени». Материалом для исследования послужили образцы ДНК, выделенные из «сухой капли» крови.

Результаты: Для изучения влияния материнских инфекционно-воспалительных осложнений беременности на количество копий TREC и KREC) обследованные новорожденные были разделены на 2 группы:

I группа ($n = 35$) – новорожденные от матерей без инфекций мочевыводящих путей и воспалительных заболеваний со сроком беременности 34,0 (32,5–34,8) недель;

II группа ($n = 65$) – новорожденные от матерей с инфекциями мочевыводящих путей и воспалительными заболеваниями при сроке беременности 34,0 (32,3–35,5) недель.

У детей от матерей без инфекционно-воспалительных заболеваний мочеполовой системы медиана копий TREC составила 25 313,0 (11 765,50–43 762,00), медиана копий KREC -16 478,0 (6 351,50–27 306, 50) соответственно. У детей от матерей с инфекционно-воспалительными заболеваниями мочеполовой системы медиана значений TREC составила 25 344,0 (11 291,0–40 495,0), копий KREC – 13 106,0 (5 323,0–22 171,0) соответственно. Статистически значимых различий в значениях копий не было. Наличие инфекционно-воспалительных заболеваний во время

беременности существенно не влияло на показатели TREC и KREC у недоношенных детей.
Таким образом, инфекционная патология репродуктивной системы матери не влияет на функциональную активность тимуса и костного мозга недоношенных детей.
Выводы: Использование определения уровней TREC и KREC позволяет определить принадлежность новорожденного ребенка, рожденного от матери с инфекционной патологией репродуктивной системы к группе риска врожденной иммунопатологии.
Финансирование: Исследование выполнено в рамках задания 02.01. «Разработать и внедрить метод отдельных нарушений, вовлекающих иммунный механизм у недоношенных новорожденных с использованием кольцевых молекул ДНК Т- и В-клеточного рецептора (TREC/KREC)» отраслевой научно-технической программы «Здоровье матери и ребенка – основа здоровья нации» на 2019–2021 годы (№ госрегистрации 20191042).

Assessment of the effect of reproductive pathology on the number of TREC/KREC copies in the dry blood spot in premature newborns

Polyakova E.[1]*, Beresten S.[2], Belevtsev M.[1]

[1] *Belarusian Research Center for Pediatric Oncology, Hematology and Immunology, Minsk, Belarus*

[2] *Republican scientific and practical center "Mother and child", Minsk, Belarus*

** polyakovakat86@gmail.com*

Key words: reproductive pathology; premature newborns; real-time PCR; TREC/KREC

Motivation and Aim: The immunological status of a newborn child is formed in the conditions of the immunological relationship between mother and fetus. Various types of reproductive pathology of the mother disrupt these relationships, which can affect not only the intrauterine development of the fetus, but also its immune system. Infectious and inflammatory complications of pregnancy increase the risk of premature birth. As a result, premature babies are born, whose immune system differs from full-term newborns in its functional immaturity. The mechanisms of changes in the innate immune system and the effect of immaturity on the overall risk of developing immunological disorders associated with a high risk of early development of infectious pathology in the neonatal period have not been fully studied, but premature birth is considered the greatest risk.
Methods and Algorithms: We studied the levels of T and B lymphocyte neogenesis markers (TREC and KREC) in 100 premature infants with a gestation period of 36-37 weeks by real-time PCR. The material for the study was DNA samples isolated from a "dry drop" of blood.
Results: To study the effect of maternal infectious and inflammatory complications of pregnancy on the number of copies of TREC and KREC), the examined newborns were divided into 2 groups: Group I (n = 35) – newborns from mothers without urinary tract

infections and inflammatory diseases with a gestation period of 34.0 (32.5–34.8) weeks; Group II (n = 65) – newborns from mothers with urinary tract infections and inflammatory diseases at a gestation period of 34.0 (32.3–35.5) weeks.
In children from mothers without infectious and inflammatory diseases of the genitourinary system, the median of TREC copies was 25 313,0 (11 765,50–43 762,00), median of KREC copies -16 478,0 (6 351,50–27 306, 50) accordingly. In children from mothers with infectious and inflammatory diseases of the genitourinary system, the median TREC values were 25 344,0 (11 291,0–40 495,0), KREC copies – 13 106,0 (5 323,0–22 171,0) accordingly. There were no statistically significant differences in the values of the copies. The presence of infectious and inflammatory diseases during pregnancy did not significantly affect the TREC and KREC indices in premature infants. Thus, the infectious pathology of the mother's reproductive system does not affect the functional activity of the thymus and bone marrow of premature infants.
Conclusion: The use of determination of TREC and KREC levels makes it possible to determine whether a newborn child born to a mother with an infectious pathology of the reproductive system belongs to the risk group of congenital immunopathology.
Funding: The study was carried out within the framework of task 02.01. "To develop and implement a method of individual disorders involving the immune mechanism in premature newborns using ring DNA molecules of the T- and B-cell receptor (TREC/KREC)" of the sectoral scientific and technical program “Maternal and child health - the basis of national health” for 2019–2021 (No. state registration 20191042).

doi 10.18699/bgrs2024-9.2-14

Каталазная активность сывороточных IgG и фрагментов IgG, элюированных с поверхности внеклеточных везикул больных с колоректальным раком: первые результаты

Смирнова Л.П.[1]*, Казанцева Д.В.[1], Воронина В.С.[2], Антипина П.А.[2], Кондакова И.В.[3], Костромицкий Д.Н.[3], Юнусова Н.В.[2, 3], Иванова С.А.[1]

[1] *Научно-исследовательский институт психического здоровья, Томский научно-исследовательский медицинский центр РАН, Томск, Россия*

[2] *Сибирский государственный медицинский университет, СибГМУ, Томск, Россия*

[3] *Научно-исследовательский институт онкологии, Томский научно-исследовательский медицинский центр РАН, Томск, Россия*

* *lpsmirnova@yandex.ru*

Ключевые слова: абзимы; колоректальный рак; экзосомы; каталазная активность

Мотивация и цель: Согласно эпидемиологическим данным, в мировой структуре КРР занимает третье место по распространённости, уступая раку легкого и раку молочной железы, а также занимает второе место по смертности от онкологических заболеваний [1]. На данный момент причины развития КРР не до конца определены, поэтому его изучение является актуальной проблемой на сегодняшний день. Известно, что окислительный стресс (ОС) является важным патогенетическим фактором развития многих заболеваний, в том числе онкологических. У больных КРР происходит нарушение в функционирование антиоксидантной системы, в результате чего происходит избыточное накопление активных форм кислорода (АФК), вызывающих окислительное повреждение клеток слизистой оболочки кишки [2]. В последние несколько лет активно изучаются IgG с оксидоредуктазными свойствами, которые были выявлены при некоторых заболеваниях [3–7]. Предполагается, что такие каталитически активные антитела (абзимы) могут компенсировать дефицит внутриклеточных антиоксидантных ферментов, потому что в отличие от тканевых ферментов, антитела присутствуют в кровотоке длительное время и в меньшей степени подвергаются протеолизу. IgG также могут переноситься будучи связанными с внеклеточными везикулами. Внеклеточные везикулы (ВВ) – это мембранно-замкнутые структуры, размером от 30 до 1000 нм, являющиеся транспортерами сложного биомолекулярного груза: белков, пептидов, липидов и нуклеиновых кислот; белки переносятся на поверхности везикул, образуя белковую корону, в состав которой входят IgG [8]. После высвобождения клетками во внеклеточное пространство ВВ транспортируют свое содержимое как к близлежащим клеткам, так и к отдаленно расположенными клетками разных тканей с помощью системы кровотока. Таким образом, ВВ участвуют в регуляции межклеточных взаимодействий здоровых тканей и опухоли. Существует исследование, показывающее увеличение популяции IgG+ВВ среди пациентов с раком поджелудочной железы; при исследовании реакции на химиотерапию, было выяснено, что происходит значительное снижение популяции IgG+ВВ у пациентов, ответивших на лечение [9]. Для элюции связанных с ВВ IgG нами

использовалась IdeZ – недавно синтезированная рекомбинантная эндопептидаза, специфически расщепляющая IgG в шарнирной области на двойной Fab-фрагмент две части Fc-фрагмента [10].

Представляется актуальным исследование каталазной активности свободных сывороточных IgG и связанных с ВВ IgG при КРР, для определения значимости вклада тех или других IgG в патогенез заболевания. Целью данного исследования является изучение каталазной активности IgG, выделенных из сыворотки крови, и фрагментов антител, элюированных с поверхности ВВ больных колоректальным раком и здоровых доноров.

Методы и алгоритмы: В исследование было включено 18 пациентов с колоректальным раком ($T_{3\text{-}4}N_{0\text{-}2}M_0$), средний возраст которых составил 64.5 [60;70] года и 16 условно-здоровых доноров (средний возраст – 55 [50.5;64.5]). Выделение малых ВВ из плазмы крови проводилось методом ультрафильтрации с двойным ультрацентрифугированием. Для изучения морфологии и размера выделенных частиц проводилась при помощи трансмиссионнной электронной томографии (Новосибирск). Распределение и концентрация выделенных везикул были определены с помощью метода анализа траекторий наночастиц. Для подтверждения экзосомальной природы выделенных везикул определяли уровень тетраспанинов CD9, CD63 и CD81 методом проточной цитометрии с укрупнением на латексных частицах, покрытых моноклональными антителами против CD9 (ab134375, Abcam). Цитометрия была выполнена на приборе Cytoflex (Beckman Coulter, США), полученные данные анализировались в программе CytExpert 2.0 Software. ВВ инкубировали с эндопротеазой IdeZ в соотношении количества белка в пробе 1:0,5. Выделение IgG из сыворотки крови и из обработанных IdeZ препаратов ВВ проводили с помощью аффинной хроматографии на колонках с Protein-G-Sepharose на хроматографе AKTA pure (GE, USA). Диализ антител проводили против 20 мМ фосфатного буфера, pH = 7.0. Концентрацию IgG определяли спектрофотометрическим методом на многофункциональном ридере Varioskan LUX (Thermo Seinctific, USA) (прибор размещен на базе ЦКП «Медицинская геномика», ТНИМЦ) при длине волны 260, 280 и 320 нм. Электрофоретический анализ полученных препаратов проводился по Леммли в градиентном 4–18 % ПААГ. Визуализация гелей проводилась с помощью системы iBright Imaging Systems FL1500 (Thermo Scientific, США) (прибор размещен на базе ЦКП «Медицинская геномика», ТНИМЦ). Каталазную активность препаратов антител определяли спектрофотометрическим методом на спектрофотометре Varian Cary 60 UV-Vis (Agilent, USA) по уменьшению оптической плотности с течением времени, вызванное разложением перекиси водорода при добавлении исследуемого образца. Статистическая обработка данных проводилась в программе Statistica 12.0.

Результаты: По соблюдению общепринятых критериев (выделение IgG методом аффинной хроматографии на аффинном сорбенте и электрофоретическая гомогенность выделенных препаратов) было доказано, что выявленная каталазная активность является собственным свойством антител. Уровень каталазной активности сывороточных IgG пациентов с колоректальным раком (995.17 ± ± 196.12 мкМ H_2O_2/мин/мг белка) достоверно ($p = 0.01$) выше, чем у здоровых лиц (445.75 ± 172.5 мкМ H_2O_2/мин/мг белка). Также выявлено, что изучаемая активность IgG зависит от стадии рака и степени дифференцировки опухоли. Каталазная активность IgG у больных с низкодифференцированным

колоректальным раком (580.94 ± 220.31 мкМ H_2O_2/мин/мг белка) была значимо (p = 0.04) ниже, чем у больных с высокодифференцированным колоректальным раком (1471.32 ± 697.7 мкМ H_2O_2/мин/мг белка). Уровень каталазной активности IgG у пациентов во второй стадии развития заболевания (494.11 ± 163.82 мкМ H_2O_2/мин/мг белка) оказался статически значимо ниже (p = 0,04), чем у пациентов с третьей стадией рака (1333.4 ± 277.69 мкМ H_2O_2/мин/мг белка).
Для получения IgG с поверхности ВВ была разработана специальная методика с использованием эндопептизы IdeZ. Препараты ВВ больных с колоректальным раком инкубировали с IdeZ после чего подвергали аффинной хроматографии на колонке с Protein-G-Sepharose. Полученные отчищенные образцы по результатам электрофореза содержали фрагменты IgG массами 20 kDa (Fc-фрагмент) и 130 kDa (Fab2-фрагмент). Полученные фрагменты IgG с поверхности ВВ также обладали каталазной активностью, однако их каталазная активность оказалась ниже, чем у соответствующих препаратов сывороточных цельных IgG пациентов с колоректальным раком.
Выводы: Проведенное исследование является пилотным в изучении абзимов с оксидоредуктазными свойствами у больных с колоректальным раком. Полученные результаты показывают перспективу дальнейшего изучения данной тематики. Было показано, что каталазная активность IgG повышается у больных с колоректальным раком, а также зависит от степени дифференцировки опухоли и стадии рака. Немаловажными являются результаты активности элюированных с поверхности ВВ фрагментов IgG. Разработана методика отчистки IgG, связанных с ВВ, также показано, что полученные препараты обладают каталазной активностью, что дает основу для дальнейшего развития нового перспективного направления исследований.

Catalase activity of serum IgG and IgG fragments eluted from the surface of extracellular vesicles from patients with colorectal cancer: a pilot study

Smirnova L.P.[1]*, Kazantseva D.V.[1], Voronina V.S. [2], Antipina P.A.[2], Kondakova I.V.[3], Kostromitsky D.N.[3],Yunusova N.V.[2, 3], Ivanova S.A.[1]

[1] *Research Institute of Mental Health, Tomsk Scientific Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia*

[2] *Siberian State Medical University, Siberian State Medical University, Tomsk, Russia*

[3] *Cancer Research Institute, Tomsk Scientific Research Medical Center of the Russian Academy of Sci-ences, Tomsk, Russia*

** lpsmirnova@yandex.ru*

Key words: abzymes; colorectal cancer; exosomes; catalase activity

Motivation and purpose: According to epidemiological data, colorectal cancer ranks third in prevalence, behind lung cancer and breast cancer, and also ranks second in mortality from cancer [1]. At the moment, the causes of colorectal cancer development are not fully determined, so its study is an urgent problem today. It is known that oxidative stress (OS) is an important pathogenetic factor in the development of many diseases, including cancer. In patients with colorectal cancer, there is a disruption in the

functioning of the antioxidant system, resulting in excessive accumulation of reactive oxygen species (ROS), causing oxidative damage to the cells of the intestinal mucosa [2]. In the last few years, IgG with oxidoreductase properties, which have been identified in some diseases, have been actively studied [3–7]. It is assumed that such catalytically active antibodies (abzymes) can compensate for the deficiency of intracellular antioxidant enzymes, because, unlike tissue enzymes, antibodies are present in the bloodstream for a long time and are less subject to proteolysis. IgG can also be transported bound to extracellular vesicles. Extracellular vesicles (EVs) are membrane-closed structures ranging in size from 30 to 1000 nm, which are transporters of complex biomolecular cargo: proteins, peptides, lipids and nucleic acids; proteins are transferred to the surface of the vesicles, forming a protein corona, which includes IgG [8]. After being released by cells into the extracellular space, EVs transport their contents both to nearby cells and to distantly located cells of different tissues via the bloodstream. Thus, EVs are involved in the regulation of intercellular interactions between healthy tissues and tumors. There is a study showing an increase in the IgG+EV population among patients with pancreatic cancer; When studying the response to chemotherapy, it was found that there was a significant decrease in the IgG+ EV population in patients who responded to treatment [9]. To elute EV-bound IgG, we used IdeZ, a recently synthesized recombinant endopeptidase that specifically cleaves IgG in the hinge region into a double Fab fragment and two parts of an Fc fragment [10].

It seems relevant to study the catalase activity of free serum IgG and EV-related IgG in CRC, to determine the significance of the contribution of one or another IgG to the overall catalase activity. The purpose of this research is to study the catalase activity of IgG isolated from blood serum and antibody fragments removed from the surface of EVs of colorectal cancer patients and healthy donors.

Methods and algorithms: The study included 18 patients with colorectal cancer (T3-4N0-2M0), whose average age was 64.5 [60;70] years and 16 apparently healthy donors (average age – 55 [50.5; 64.5]). Isolation of small explosives from blood plasma was carried out by ultrafiltration with double ultracentrifugation. To study the morphology and size of the isolated particles, it was carried out using transmission electron tomography (Novosibirsk). The distribution and concentration of the isolated vesicles were determined using the nanoparticle trajectory analysis method. To confirm the exosomal nature of the isolated vesicles, the level of tetraspanins CD9, CD63 and CD81 was determined by flow cytometry with enlargement on latex particles coated with monoclonal antibodies against CD9 (ab134375, Abcam). Cytometry was performed on a Cytoflex instrument (Beckman Coulter, USA), and the data obtained were analyzed in CytExpert 2.0 Software. EVs were incubated with IdeZ endoprotease in a ratio of the amount of protein in the sample of 1:0.5. Isolation of IgG from blood serum and from IdeZ-treated IV preparations was carried out using affinity chromatography on Protein-G-Sepharose columns on an AKTA pure chromatograph (GE, USA). Antibodies were dialyzed against 20 mM phosphate buffer, pH = 7.0. The IgG concentration was determined by the spectrophotometric method on a multifunctional Varioskan LUX reader (Thermo Seinctific, USA) (the device is located at the Medical Genomics Center for Common Use, TNIMC) at wavelengths of 260, 280 and 320 nm. Electrophoretic analysis of the obtained preparations was carried out according to Laemmli in a gradient of 4–18 % PAGE. The gels were visualized using the bright Imaging Systems FL1500 (Thermo Scientific, USA) (the device is located at the Medical Genomics Center for Common Use, TNIMC).The catalase activity of antibody preparations was determined

spectrophotometrically on a Varian Cary 60 UV-Vis spectrophotometer (Agilent, USA) by the decrease in optical density over time caused by the decomposition of hydrogen peroxide upon addition of the test sample. Statistical data processing was carried out using the Statistica 12.0 program.

Results: Based on compliance with a number of generally accepted criteria (isolation of IgG by affinity chromatography on an affinity sorbent and electrophoretic homogeneity of the isolated preparations), it was proven that the identified catalase activity is an intrinsic property of antibodies. The level of catalase activity of serum IgG from patients with colorectal cancer (995.17 ± 196.12 μM H_2O_2/min/mg protein) was significantly ($p = 0.01$) higher than in healthy individuals (445.75 ± 172.5 μM H_2O_2/min/mg protein). It was also found that the studied IgG activity depends on the stage of cancer and the degree of tumor differentiation. Catalase activity of IgG in patients with poorly differentiated colorectal cancer 580.94 ± 220.31 μM H_2O_2/min/mg protein was significantly ($p = 0.04$) lower than in patients with well-differentiated colorectal cancer 1471.32±697.7 μM H_2O_2/ min/mg protein. The level of catalase activity of IgG in patients in the second stage of the disease (494.11±163.82 μM H_2O_2/min/mg protein) was statistically significantly lower ($p = 0.04$) than in patients with the third stage of cancer (1333.4 ±277.69 μM H_2O_2/min/mg protein). To obtain IgG from the surface of EVs, a special technique was developed using endopeptyse IdeZ. IV preparations from patients with colorectal cancer were incubated with IdeZ and then subjected to affinity chromatography with a Protein-G-Sepharose column. According to the results of electrophoresis, the resulting purified samples contained IgG fragments with masses of 20 kDa (Fc fragment) and 130 kDa (Fab2 fragment). The resulting IgG fragments from the EV surface also had catalase activity, but their catalase activity was lower than that of the corresponding whole IgG serum preparations from patients with colorectal cancer.

Conclusions: The study is a pilot study of abzymes with oxidoreductase properties in patients with colorectal cancer. The results obtained show prospects for further study of this topic. It has been shown that IgG catalase activity is increased in patients with colorectal cancer and also depends on the degree of tumor differentiation and the stage of cancer. The results of the activity of IgG fragments eluted from the EVs surface are also important. A method for purifying IgG associated with EVs has been developed, and it has also been shown that the resulting preparations have catalase activity, which provides the basis for the further development of a new promising area of research.

Список литературы/References

1. Sung H., Ferlay J., Siegel R.L., Laversanne M., Soerjomataram I., Jemal A., Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi 10.3322/caac.21660
2. Bardelčíková A., Šoltys J., Mojžiš J. Oxidative stress, inflammation and colorectal cancer: An overview. *Antioxidants*. 2023;12(4):901. doi 10.3390/antiox12040901
3. Ermakov E.A., Smirnova L.P., Bokhan N.A., Semke A.V., Ivanova S.A., Buneva V.N., Nevinsky G.A. Catalase activity of IgG antibodies from the sera of healthy donors and patients with schizophrenia. *PLoS One*. 2017;12(9):e0183867. doi 10.1371/journal.pone.0183867
4. Tolmacheva A.S., Blinova E.A., Ermakov E.A., Buneva V.N., Vasilenko N.L., Nevinsky G.A. IgG abzymes with peroxidase and oxidoreductase activities from the sera of healthy humans. *J Mol Recognit*. 2015;28(9):565-580. doi 10.1002/jmr.2474
5. Smirnova L.P., Mednova I.A., Krotenko N.M., Alifirova V.M., Ivanova S.A. IgG-dependent dismutation of superoxide in patients with different types of multiple sclerosis and healthy subjects. *Oxid Med Cell Longev*. 2020;2020:8171020. doi 10.1155/2020/8171020
6. Епимахова Е.В., Смирнова Л.П., Казанцева Д.В., Паршукова Д.А., Кротенко Н.М., Васильева А.Р., Иванова С.А., Семке А.В. Оценка цитотоксических эффектов IgG, выделенных из

сыворотки крови больных шизофренией. *Сибирский вестник психиатрии и наркологии*. 2021;4(113):5-13. doi 10.26617/1810-3111-2021-4(113)-5-13
[Epimakhova E.V., Smirnova L.P., Kazantseva D.V., Parshukova D.A., Krotenko N.M., Vasilyeva A.R., Ivanova S.A., Semke A.V. Evaluation of the cytotoxic effects of IgG isolated from the blood serum of patients with schizophrenia. *Siberian Bulletin of Psychiatry and Narcology*. 2021;4(113):5-13 (in Russian)]
7. Ермаков Е.А., Смирнова Л.П., Кротенко Н.М., Семке А.В., Бунева В.Н., Невинский Г.А. Каталазная активность каталитических антител при шизофрении. *Российский иммунологический журнал*. 2019;13(2-1):242-244. doi 10.31857/S102872210006588-1
[Ermakov E.A., Smirnova L.P., Krotenko N.M., Semke A.V., Buneva V.N., Nevinsky G.A. Catalase activity of catalytic antibodies in schizophrenia. *Russian Immunological Journal*. 2019;13(2-1):242-244 (in Russian)]
8. Peterson M.F., Otoc N., Sethi J.K., Gupta A., Antes T.J. Integrated systems for exosome investigation. *Methods*. 2015;87:31-45. doi 10.1016/j.ymeth.2015.04.015
9. Couto N., Elzanowska J., Maia J., Batista S., Pereira C.E., Beck H.C., Carvalho A.S., Strano Moraes M.C., Carvalho C., Oliveira M., Matthiesen R., Costa-Silva B. IgG+ extracellular vesicles measure therapeutic response in advanced pancreatic cancer. *Cells*. 2022;11(18):2800. doi 10.3390/cells11182800
10. Бокша И.С. Рекомбинантные эндопептидазы IdeS и IdeZ и возможный потенциал их применения. *Биохимия*. 2023;6(88):900-912. doi 10.31857/S0320972523060027
[Boksha I.S. Recombinant endopeptidases Des and Idea and the possible potential of their application. *Biochemistry*. 2023;6(88):900-912 (in Russian)]

doi 10.18699/bgrs2024-9.2-15

Идентификация бактерий, способствующих улучшению результатов иммунотерапии меланомы

Строкач А.А.[1]*, Морозов М.Д.[1], Зорук П.Ю.[1], Веселовский В.А.[1], Колдман В.А.[1, 2], Колдман С.Д.[1, 2], Олехнович Е.И.[1], Климина К.М.[1]

[1] *Федеральный научно-клинический центр физико-химической медицины имени академика Ю.М. Лопухина Федерального медико-биологического агентства России, Москва, Россия*

[2] *Государственный научный центр Федеральный медицинский биофизический центр им. А.И. Бурназяна Федерального медико-биологического агентства России, Москва, Россия*

* *alexandra.vlasova.2017@yandex.ru*

Ключевые слова: иммунотерапия анти-PD1; NGS; кишечная микробиота

Мотивация и цель: Микробиота кишечника человека выполняет множество функций важных для поддержания здоровья и регуляции иммунной системы. В научной литературе имеется связь между составом микробиоты кишечника и результатами иммунотерапии злокачественных опухолей [1]. Эта зависимость была подтверждена экспериментами по пересадке микробиоты от пациентов мышам, лишенным микробиоты, а также пациентам с негативными результатами лечения [2, 3], однако четкого понимания биологических механизмов все еще нет. Поэтому целью данного исследования была систематизация метагеномных данных из открытых источников и валидирование полученных результатов с использованием экспериментальной модели меланомы.

Методы и алгоритмы: С использованием модели смешанных эффектов (ZicoSeq) были проанализированы таксономические профили 680 кишечных метагеномов пациентов, проходивших иммунотерапию, из семи ранее опубликованных исследований (базы данных NCBI-EBI). Для повышения достоверности результатов в анализе были использованы различные стратегии коррекции эффектов множественного тестирования гипотез. В частности, использовались поправки на множественное сравнение: FWER (family-wise error rate) – вероятность совершить хотя бы одну ошибку первого рода и FDR (false discovery rate) – для учета вероятности ошибки второго рода. Для проведения исследования на мышиной модели меланомы был выбран штамм B. adolescentis 150 [4]. В исследованиях использовались самки мышей C57Bl/6 весом 17–20 г, приобретенные в питомнике «Филиал Столбовая ФГБУН НЦБМТ ФМБА России». Мышам была привита культура клеток меланомы B16/F1. Вводилась лиофильная культура бифидобактерий и раствор антитела PD-1 в соответствии с разработанным протоколом исследования. В дальнейшем проводился сбор мышиных фекалий для оценки бактериального разнообразия посредством секвенирования гена 16S рРНК на Oxford Nanopore Technologies. Транскриптомное севенирование тотальной РНК опухоли применялось для количественной оценки уровня экспрессии генов.

Результаты: В результате анализа *Bifidobacterium adolescentis* была идентифицирована как единственный строгий биомаркер положительного исхода иммунотерапии, прошедший установленные пороги достоверности (FWER $p < 0.05$; FDR $p < 0.05$). Чтобы подтвердить потенциальное воздействия

B. adolescentis на результаты иммунотерапии было проведено исследование влияния анти-PD-1 терапии в комбинации с бифидобактериями на лечение экспериментальной модели меланомы.
Выводы: Результаты экспериментов на мышах показали, что введение штамма B. adolescentis 150 не только способствовал увеличению роста экспериментальной меланомы, но также приводил к повышению количества условно-патогенных бактерий в кишечнике. На основе анализа данных транскриптома, повышалась экспрессия опухолевых генов Ero1l, Mtss1l, Atf3, Mitf, Mxi1, Vegfa, Slc2a1, которые, согласно литературным данным, ассоциируются с метастазированием и прогрессированием опухоли [5–7].
Финансирование: Работа выполнена при поддержке гранта Российского Научного Фонда, соглашение № 22-75-10029. Информация о проекте доступна по ссылке https://rscf.ru/project/22-75-10029/.

Identification of bacteria contributing to the improvement of melanoma immunotherapy results

Strokach A.[1]*, Morozov M.[1], Zoruk P.[1], Veselovsky V.[1], Koldman V.[1, 2], Koldman S.[1, 2], Olekhnovich E.[1], Klimina K.[1]

[1] *Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine, Moscow, Russia*
[2] *Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russia*
** alexandra.vlasova.2017@yandex.ru*

Key words: immunotherapy anti-PD1; NGS; gut microbiota

Motivation and Aim: The human gut microbiome plays an important role in maintaining health and regulating the immune system. There is a growing body of scientific evidence linking the composition of intestinal microbiota to the effectiveness of immunotherapy for cancer [1]. However, the exact mechanisms behind this relationship are still not fully understood.
This study aims to systematically analyze metagenomic data from publicly available sources and validate its findings using an experimental model of melanoma.
Methods and Algorithms: B. adolescentis 150 was selected for the study in a mouse model of melanoma [4]. The studies used female C57BL/6 mice, weighing between 17 and 20 g, which were purchased from the "Stolbovaya Branch of the Federal State Budgetary Educational Institution of the National Research Institute of FMBA of Russia." The mice were then vaccinated with a melanoma cell culture derived from the B16/F1 cell line. After that, a lyophilized culture of bifidobacteria and a PD-1 antibody solution were injected into the mice according to the developed research protocol. Later, faecal samples were collected from the mice to assess bacterial diversity using sequencing of the 16S rRNA gene at Oxford Nanopore Technologies. In addition, transcriptomic sequencing of the total tumor RNA was performed to quantify gene expression levels.
Results: As a result of the analysis, Bifidobacterium adolescentis was identified as the only strict biomarker for a positive outcome from immunotherapy that met the established confidence thresholds (FWER $p < 0.05$; FDR $p < 0.05$).

To confirm the potential impact of B. adolescentis on immunotherapy outcomes, a study was conducted to investigate the effect of anti-PD-1 therapy combined with *B. adolescentis* in an experimental melanoma model.

Conclusion: The results from experiments on mice demonstrated that the administration of the *B. adolescentis* 150 strain not only promoted the growth of experimental melanoma but also increased the number of opportunistic bacteria in the intestines. Based on transcriptome data analysis, the expression levels of genes associated with tumor progression and metastasis, such as Ero1l, Mtss1l, Atf3, Mitf, Mxi1, and Vegfa, increased. These genes have been linked to tumor metastasis and disease progression in previous studies [5–7].

Funding: Financial support for this study was provided by the Russian Science Foundation under the grant number 22-75-10029, available at https://rscf.ru/project/22-75-10029/.

Список литературы/References

1. Lu Y. et al. Gut microbiota influence immunotherapy responses: mechanisms and therapeutic strategies. *J Hematol Oncol*. 2022;15(1):47. doi 10.1186/s13045-022-01273-9
2. Routy B. et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science*. 2018;359:91-97
3. Gopalakrishnan V. et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science*. 2018;359:97-103
4. Dyachkova M.S. et al. Draft genome sequences of bifidobacterium angulatum GT102 and bifidobacterium adolescentis 150: focusing on the genes potentially involved in the gut-brain axis. *Genome Announc*. 2015;3(4):e00709-15. doi 10.1128/genomeA.00709-15
5. Kajiwara T. et al. Hypoxia augments MHC class I antigen presentation via facilitation of ERO1-α-mediated oxidative folding in murine tumor cells. *Eur J Immunol*. 2016;46:2842-2851
6. Muto S. et al. Wnt/β-Catenin signaling and resistance to immune checkpoint inhibitors: from non-small-cell lung cancer to other cancers. *Biomedicines*. 2023;11(1):190. doi 10.3390/biomedicines11010190
7. Rodolfo M., Daniotti M., Vallacchi V. Genetic progression of metastatic melanoma. *Cancer Lett*. 2004;214:133-147

doi 10.18699/bgrs2024-9.2-16

Поиск и изучение противоинвазивного действия ряда производных терпеноидов и прочих природных соединений на клетки перевиваемых линий глиобластомы

Усенов К.[1, 3]*, Покровский А.Г.[1], Чересиз С.В.[1], Хамад М.С.[1], Покровский М.А.[3], Ковалева К.С.[2], Яровая О.И.[2], Салахутдинов Н.Ф.[2]

[1] *Новосибирский государственный университет, Новосибирск, Россия*

[2] *Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия*

[3] *Лаборатория молекулярной патологии ИМПЗ НГУ, Новосибирск, Россия*

* *kubanych.usenov@mail.ru*

Ключевые слова: Глиобластома; инвазия; миграция; тест царапина; онкология

Мотивация и цель: Глиобластома (ГБМ) представляет собой наиболее часто встречающуюся и обладающую высокой степенью злокачественностью среди первичных опухолей мозга у взрослых, агрессивность которой в значительной степени определяется ее способностью к инвазии – активной инфильтрации отдельных злокачественных клеток или их групп в окружающую ткань головного мозга. Современный стандарт лечения для пациентов с глиобластомой включает по возможности наиболее радикальную хирургическую резекцию опухоли с последующей лучевой терапией и адъювантной химиотерапией темозоломидом, однако этот режим не обеспечивает длительной выживаемости пациентов (медиана которой составляет 12–15 мес с момента постановки диагноза, а пятилетняя выживаемость пациентов с ГБМ не превышает 5 %). Как следует из сравнения значений медианы выживаемости пациентов, в режим которых включалась химиотерапия, по сравнению с этим же показателем пациентов, ее не получивших (14,6 и 12,2 мес соответственно), эффективность химиотерапии к глиобластоме на сегодняшний день является невысокой. В то же время известно, что более чем у половины прооперированных пациентов с ГБМ, достаточно быстро возникает рецидив опухоли непосредственно на месте расположения послеоперационной полости или в ее краевой зоне (не более 2 см от полости), а примерно у 20 % пациентов развиваются удаленные либо множественные рецидивы опухоли. Рецидивы ГБМ, являющиеся ключевой причиной низкой выживаемости прооперированных пациентов, связывают это с процессами инвазии, в первом случае – обратной инвазии (реинвазии) клеток ГБМ, уже мигрировавших из первичной опухоли, обратно в зону послеоперационной полости, во втором – инвазией в удаленные от первичной опухоли участки мозга. Молекулярные и клеточные механизмы инвазии ГБМ сейчас активно исследуются *in vitro* и в животных моделях, однако в клинической практике подходы к химиотерапии ГБМ, основанные на подавлении инвазии клеток этой опухоли, на сегодняшний день отсутствуют.

Цель работы: Поиск и изучение новых потенциальных химиотерапевтических препаратов на основе производных терпеноидов, обладающих протиинвазивной активностью в отношении клеток глиобластомы.

Методы и алгоритмы: Клеточные линии. Перевиваемые клеточные линии глиобластомы SNB-19 (теперь известна под названием U251-MG). Коллекция производных терпеноидов и прочих природных соединений, исследуемая в работе, предоставлена сотрудниками Новосибирского института органической химии им. Н.Н. Ворожцова СО РАН. Коллекция состоит из 93 оригинальных соединений, включающих в себя несколько серий. *KS* серия – дитерпеноиды, производные смоляных кислот. Дизайн эксперимента:

1. Определение токсичности веществ G15 % в которых вещества будут изучаться в тесте царапины. Жизнеспособность клеток для определения диапазонов токсичности. Выполнено методикой МТТ согласно протоколу.
2. Анализ инвазии методом заживления царапины/раны (wound healing assay/scratch assay). Данные были получены для клеток SNB-19 выращенных на пластиковых 6 и 24-луночных планшетах (2 млн и 300 000 клеток/лунку) до достижения 100 % плотности конфлуэнтности клеточного монослоя, после чего стандартная царапина шириной 700 мкм ориентированная вертикально относительно лунки наносилась кончиком в конфлуэнтном монослое.

Результаты: Изучена цитотоксичность исследуемых соединений для определения цитотоксического диапазона концентраций для использования в тесте зарастания царапины клеточной линии. Результаты экспериментов задокументировали и обрабатывали (производили измерение ширины царапины) с помощью системы анализа изображений клеток Paula (Leica). Изображения были получены в моменты времени 0 и 24 ч с использованием объектива Neofuoar 0,3 (10×) и длины волны источника света 633 нм. При анализе наших результатов анализа заживления ран мы рассчитали среднюю ширину царапины в каждой лунке через 0, и 24 ч. Затем мы рассчитали среднюю скорость закрытия, разделив среднюю ширину царапины на время. Затем мы нормализовали полученные числа к средней скорости закрытия нашего контроля по формуле Vотн KS = (17,5:30,70) x 100 % = 57 %. Сравнивая результаты, мы пришли к выводу, что потенциально активные антиинвазивные вещества из 93 исследованных соединений, которые имели более низкие цитотоксические концентрации, их влияние на инвазию клеток SNB-19 ГБМ анализировали методом закрытия царапин. Идентифицировано одно (KS 481) соединение с высокой антиинвазивной активностью.

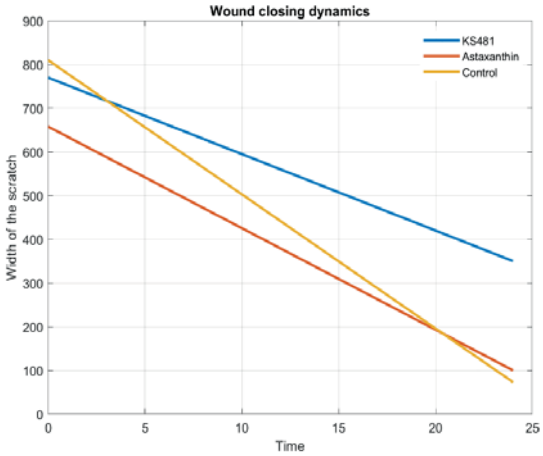


Рис. 1. Скорость закрытия царапины по сравнению с астаксантином и контролем

	Название	GI15% (μM)	GI50% (μM)
Высокотоксичные соединения	KS408	>1	5.731
	KS524	>1	12.556
	KS214	>1	17.9
	KS426	>1	2.244
	KS482	>1	10.753
	KS406	>1	21.65
	KS514	>1	14.769
	KS521	>1	8.345
	KS482	>1	12.766
	KS467	>1	27.68
	KS507	>1	11.937
Низкотоксичные соединения	KS522	14.655	86.13
	KS471	9.85	54.877
	KS481	12.315	76.433
	KS475	6.323	53.143
	KS508	58.827	146.533
	KS520	41.626	122.666
	KS496	19.912	94.353
	KS450	7.577	59.03
	KS452	29.752	104.593
	KS470	4.064	42.481
	KS480	3.772	37.433

Таблица 1. Результаты МТТ-теста цитотоксичности 11 низкотоксичных веществ из 93. Они были выбраны для анализа заживления ран

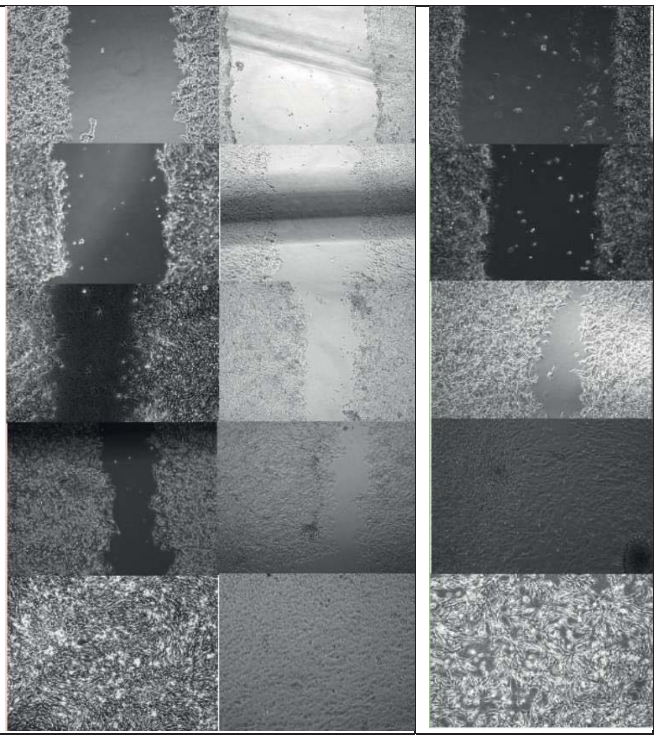


Рис. 2. Изображения анализа заживления ран на 0 и 24 ч с веществом а) KS 481, б) астаксантин, в) ДМСО

Выводы: В ходе исследования было изучено на цитотоксичность 93 производных терпеноидов на клеточных линиях глиобластомы SNB19, что позволило выделить группу из 11 соединений с низкой токсичностью в отношении клеток глиобластомы. На следующем этапе, с помощью теста зарастания царапины/раны (wound-healing/scratch test) была исследована потенциальная противоинвазивная активность отобранных соединений в отношении клеток глиобластомы. Наиболее активное вещество (KS481) продемонстрировал значительный эффект подавления инвазии, снижая скорость зарастания царапин на 57 % в клеточной линии SNB 19. Для улучшения ингибирования вещества планируется оптимизации структуру молекулы совместно с институтом органической химии.

Финансирование: Исследование поддержано FSUS № 2020-0035.

Search and study of anti-invasive action of a number of derivatives of terpenoids and other natural compounds on cells of transplanted glioblastoma lines

Usenov K.[1, 3]*, Pokrovsky A.G.[1], Cheresiz S.V.[1], Hamad M.C.[1], Pokrovsky M.A.[3], Kovaleva K.[2], Yarovaya O.I.[3], Salakhutdinov N.F.[2]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *N.N. Vorozhtsov Institute of Organic Chemistry SB RAS, Novosibirsk, Russia*

3 Laboratory of Molecular Pathology IMPZ, Novosibirsk State University, Novosibirsk, Russia

* *kubanych.usenov@mail.ru*

Key words: Glioblastoma, invasion, migration, scratch test, oncology

Motivation and Aim: Glioblastoma (GBM) is the most common and highly malignant of primary brain tumours in adults. Its aggressiveness is largely determined by its ability to invade – the active infiltration of individual malignant cells or groups of cells into the surrounding brain tissue. The modern standard of treatment for patients with glioblastoma includes the most radical surgical resection of the tumor with subsequent radiotherapy and adjuvant chemotherapy with temozolomide, but this regimen does not ensure long-term survival of patients (median survival is 12–15 months from the moment of diagnosis, and 5-year survival rate of patients with GBM does not exceed 5 %). As can be seen from the comparison of median survival of patients whose regimen included chemotherapy compared to that of patients who did not receive it (14.6 and 12.2 months respectively), the efficacy of chemotherapy for glioblastoma is currently low. At the same time, it is known that more than half of operated patients with GBM, rather quickly develop tumor recurrence directly at the site of the postoperative cavity or in its marginal zone (not more than 2 cm from the cavity), and about 20 % of patients develop distant or multiple tumor recurrences. Recurrences of GBM, which is a key cause of low survival rate of operated patients, are attributed to the processes of invasion, in the first case – back invasion (reinvasion) of GBM cells, which have already migrated from the primary tumor, back into the area of the postoperative cavity, in the second case – invasion into the areas of the brain distant from the primary tumor. Molecular and cellular mechanisms of GBM invasion are now actively investigated in vitro and in animal models, however, in clinical practice, approaches to GBM chemotherapy based on suppression of invasion of cells of this tumor are currently absent.

Purpose of the work. Search and study of new potential chemotherapeutic agents based on terpenoid derivatives with anti-invasive activity against glioblastoma cells.

Methods and Algorithms: Methods and algorithms: Cell lines. Transgenic glioblastoma cell lines SNB-19 (now known as U251-MG). The collection of derivatives of terpenoids and other natural compounds investigated in this work was provided by the staff of the N.N.Vorozhtsov Institute of Organic Chemistry. N.N.Vorozhtsov Institute of Organic Chemistry SB RAS. The collection consists of 93 original compounds including several series. KS series – diterpenoids, derivatives of resin acids. Experiment design: 1. Determination of toxicity of substances G15 % in which substances will be studied in scratch test. Cell viability to determine the ranges of toxicity. Performed by MTT technique according to the protocol.

2. Analysis of invasion by scratch/scratch assay (wound healing assay/scratch assay). Data were obtained for SNB-19 cells grown on plastic 6 and 24-well plates (2.0 ml – 300,000 cells/well) until 100 % confluent cell monolayer density was achieved, after which a standard 700 µm wide scratch oriented vertically relative to the well was applied with a tip in the confluent monolayer.

Results: The cytotoxicity of the tested compounds was studied to determine the cytotoxic concentration range for use in the cell line scratch overgrowth test. Experimental results were documented and processed (scratch width measurement) using a Paula cell image analysis system (Leica). Images were acquired at time points 0 and 24 hours using a Neofuoar 0.3 (10×) objective and a light source wavelength of 633 nm. When analysing our wound healing assay results, we calculated the mean scratch width in each well after 0, and 24 hours. We then calculated the average closure rate by dividing the average scratch width by time. We then normalised the numbers obtained to the average closure rate of our control using the formula by Votn KS = (17.5:30.70) x 100 %=57 %. Comparing the results, we concluded that the potentially active anti-invasive substances among the 93 compounds tested, which had lower cytotoxic concentrations, their effects on invasion of SNB-19 GBM cells were analysed by scratch closure method. One (KS 481) compound with high anti-invasive activity was identified.

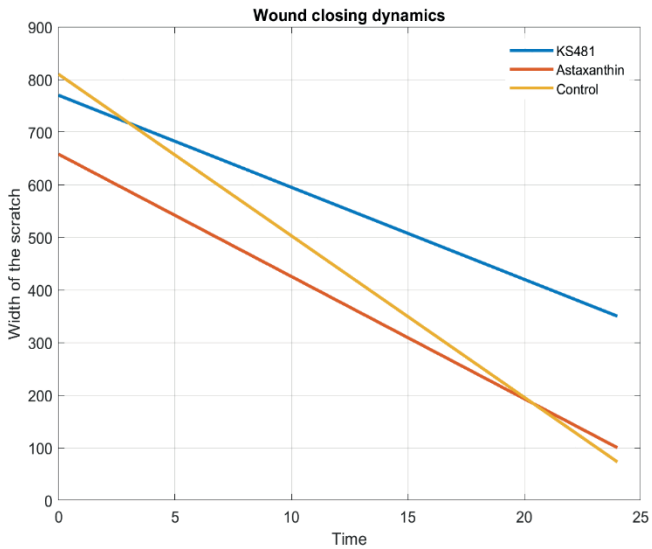


Fig. 1. Scratch closure rate compared with astaxanthin and control

	Название	GI15% (μM)	GI50% (μM)
Высокотоксичные соединения	KS408	>1	5.731
	KS524	>1	12.556
	KS214	>1	17.9
	KS426	>1	2.244
	KS482	>1	10.753
	KS406	>1	21.65
	KS514	>1	14.769
	KS521	>1	8.345
	KS482	>1	12.766
	KS467	>1	27.68
	KS507	>1	11.937
Низкотоксичные соединения	KS522	14.655	86.13
	KS471	9.85	54.877
	KS481	12.315	76.433
	KS475	6.323	53.143
	KS508	58.827	146.533
	KS520	41.626	122.666
	KS496	19.912	94.353
	KS450	7.577	59.03
	KS452	29.752	104.593
	KS470	4.064	42.481
	KS480	3.772	37.433

Table 1. Results of MTT cytotoxicity test of 11 low toxic substances out of 93. They were selected to analyze wound healing

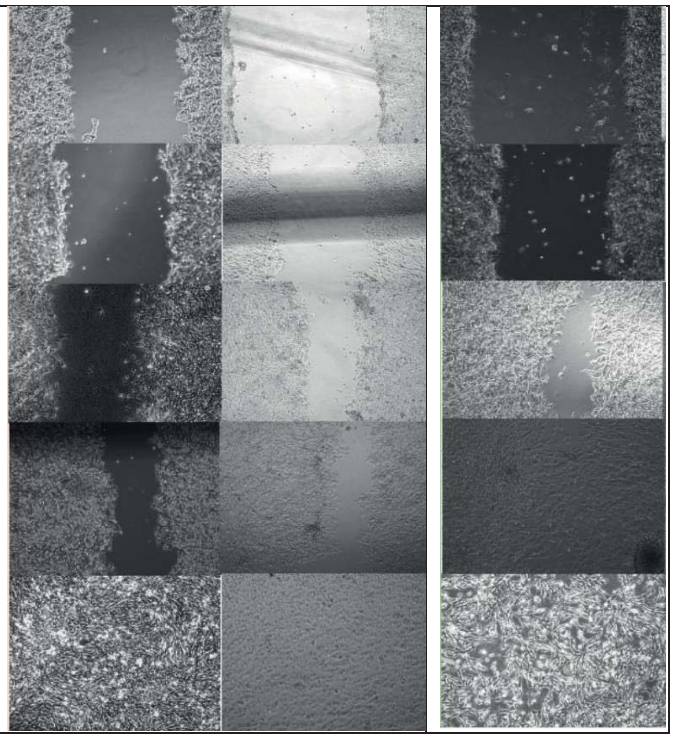


Fig. 2. Images of wound healing analysis at 0 and 24 hours with substance a) KS 481, b) astaxanthin, c) DMSO

Conclusions: In this study, 93 terpenoid derivatives were examined for cytotoxicity on SNB19 glioblastoma cell lines, identifying a group of 11 compounds with low toxicity against glioblastoma cells. In the next step, the potential anti-invasive activity of the selected compounds against glioblastoma cells was investigated using the wound-healing/scratch test. The most active compound (KS481) showed a significant invasion suppression effect, reducing the rate of scratch overgrowth by 57% in the cell line SNB 19. To improve the inhibition of the compound it is planned to optimize the structure of the molecule in collaboration with the Institute of Organic Chemistry.

Funding: The research was supported by FSUS No. 2020-0035.

Список литературы/Reference

1. Vollmann-Zwerenz A., Leidgens V., Feliciello G., Klein C.A., Hau P. Tumor Cell Invasion in Glioblastoma. *Int J Mol Sci.* 2020;21(6):1932
2. Kramer N. et al. In vitro cell migration and invasion assays. *Mutat Res*. 2013;752(1):10-24. doi 10.1016/j.mrrev.2012.08.001
3. Tsuji S. et al. Antitumour Effects of Astaxanthin and Adonixanthin on Glioblastoma. *Mar Drugs*. 2020;18(9):474. doi 10.3390/md18090474

doi 10.18699/bgrs2024-9.2-17

Высокочувствительный микроанализ на основе искусственной люциферазы NLuc для диагностики социально значимых заболеваний

Франк Л.А.*, Красицкая В.В., Башмакова Е.Е., Кудрявцев Н.А., Панамарев Н.С.
Институт биофизики СО РАН, ФИЦ КНЦ СО РАН, Красноярск, Россия
** lfrank@yandex.ru*

Ключевые слова: люцифераза NLuc; биолюминесцентный микроанализ

Люцифераза NanoLuc – искусственный фермент, полученный в результате мутагенеза каталитической субъединицы целентеразин-зависимой люциферазы глубоководной креветки *Oplophorus gracilirostris*. Малый молекулярный вес (19 кДа), уникально яркая биолюминесценция, длительная кинетика сигнала (($t_{1/2}$ = 2.5 ч) и высокая термостабильность определили перспективность использования этого белка в качестве репортера в различных аналитических системах [1]. Биоспецифические соединения NanoLuc получают с помощью генетического фьюзинга, поскольку, как и другие люциферазы, фермент теряет активность при химическом конъюгировании. Это существенно ограничивает разнообразие биолюминесцентных репортеров на основе NanoLuc. Целью данного исследования было разработать способ получения сайт-специфичных высокоактивных конъюгатов этого фермента с разнообразными молекулами и тестировать полученные конъюгаты в молекулярном микроанализе ряда диагностически важных мишеней.

Методами генетической инженерии разработан дизайн и сконструирована плазмида, содержащая ген кодирующий вариант люциферазы NanoLuc(C164S)-LCTPSR (NLuc) с уникальным остатком цистеина на С-конце. Соответствующий белок получен бактериальным синтезом, очищен хроматографически и изучены его свойства. Показано, что по характеристикам биолюминесценции и стабильности полученный белок не отличается от исходной люциферазы NanoLuc. Обнаружено, что биолюминесцентный сигнал NLuc инициируется различными субстратами – фуримазином (производное целентеразина), целентеразином, и их стабильными водорастворимыми комплексами с Ca^{2+}-зависимым целентеразин-связывающим белком *Renilla* [2]. Разработан способ синтеза конъюгатов модифицированной люциферазы с различными соединениями – биотином, антителами, олигонуклеотидами и показаны их преимущества как репортеров в высокочувствительном микроанализе на основе аналогичных производных Ca^{2+}-регулируемого фотопротеина обелина [3].

Конъюгаты с антителом против кардиального тропонина I, а также с ДНК-аптамером к кардиальному белку, связывающему жирные кислоты (кБСЖК) были испытаны как репортеры в микроанализе соответствующих кардиомаркеров в модельных образцах сыворотки человека. Показана линейная зависимость сигнала в диагностически важном диапазоне концентрации мишеней с пределом обнаружения 5.4 и 1.2 нг/мл для кБСЖК и тропонина I соответственно.

Конъюгаты люциферазы с поликлональными антителами мыши против онкомаркера сурвивина были испытаны в анализе этого белка в модельных и клинических образцах мочи, и полученные результаты удовлетворительно совпадали с таковыми коммерческого колориметрического набора для ИФА (r = 0,814, p < 0,0001).

Разработан твердофазный биолюминесцентный микроанализ онкомаркера MIA (белок с меланома-ингибирующей активностью) на основе полученных нами ДНК-аптамеров и люциферазы NLuc. Получены успешные результаты тестирования способа на примере модельных стандартных и контрольных сыворотках человека.

Получены конъюгаты миниантител мыши к вирусу клещевого энцефалита с фрагментами NLuc, способными к комплементации с восстановлением биолюминесцентной активности. Показана принципиальная возможность обнаружения ВКЭ в модельных образцах и экстрактах клещей с помощью однофазного биолюминесцентного анализа.

Разработанный вариант люциферазы NLuc обладает свойствами, близкими таковым интактной люциферазы NanoLuc и не теряет биолюминесцентную активность при химическом конъюгировании с различными биоспецифическими молекулами. Полученные на ее основе конъюгаты тестированы в молекулярном микроанализе ряда диагностически важных мишеней (кардиальных и онкомаркеров, а также корпускулярного антигена) и показана перспективность их использования для разработки тест-систем на их основе.

Финансирование: Исследование выполнено в рамках ГЗ Министерства науки и высшего образования РФ (проект № FWES-2022-0002).

Highly sensitive microassay based on artificial luciferase NLuc for the diagnosis of socially significant diseases

Frank L.A.*, Krasitskaya V.V., Bashmakova E.E., Kudryavtsev N.A., Panamarev N.S.

Institute of Biophysics, FRC "Krasnoyarsk Science Center SB RAS", Krasnoyarsk, Russia

** lfrank@yandex.ru*

Key words: NLuc luciferase; bioluminescence microassay

NanoLuc luciferase is an artificial enzyme obtained by mutagenesis of the catalytic subunit of coelenterazine-dependent luciferase from the deep-sea shrimp *Oplophorus gracilirostris*. Low molecular weight (19 kDa), uniquely bright bioluminescence, long signal kinetics (($t_{1/2}$ = 2.5 h) and high thermal stability determined the prospects for using this protein as a reporter in various analytical systems [1]. NanoLuc biospecific compounds are obtained with the use of genetic fusing, since, like other luciferases, the enzyme loses activity upon chemical conjugation. This significantly limits the diversity of NanoLuc-based bioluminescent reporters. The goal of this study was to develop a method for producing site-specific, highly active conjugates of this enzyme with a variety of molecules and to test the resulting conjugates as labels in molecular microanalysis of diagnostically important targets.

Using genetic engineering methods, a plasmid was constructed containing the gene encoding the luciferase variant NanoLuc(C164S)-LCTPSR (NLuc) with a unique cysteine residue at the C-terminus. The corresponding protein was obtained by bacterial

synthesis, purified by chromatography, and its properties were studied. According to the characteristics of bioluminescence and stability, the resulting protein does not differ from intact NanoLuc. It was found that the bioluminescent signal of NLuc is initiated by various substrates – furimazine (a coelenterazine derivative), coelenterazine, and their stable water-soluble complexes with the Ca^{2+}-dependent coelenterazine-binding protein *Renilla* [2]. A method was developed for the synthesis of modified luciferase conjugates with various compounds – biotin, antibodies, oligonucleotides – and their advantages as reporters in highly sensitive microanassay compared to the similar derivatives of the Ca^{2+}-regulated photoprotein obelin was demonstrated [3].

Conjugates with an antibody against the cardiac biomarker troponin I, as well as with a DNA aptamer against heart fatty acid binding protein (hFABP) were tested as reporters in microassays of relevant cardiac markers in model human serum samples. A linear dependence of the signal is shown in the diagnostically important range of target concentrations with a detection limit of 5.4 and 1.2 ng/mL for hFABP and troponin I respectively.

Luciferase conjugates with mouse polyclonal antibodies against the tumor marker survivin were tested in the analysis of this protein in model and clinical urine samples, and the results obtained satisfactorily coincided with those of a commercial colorimetric ELISA kit ($r = 0.814$, $p < 0.0001$).

A solid-phase bioluminescent microassay for the tumor marker MIA (a protein with melanoma-inhibitory activity) has been developed based on our unique anti-MIA DNA aptamers and NLuc luciferase. Successful results of testing the method were obtained using model standard and control human sera.

Conjugates of mouse mini-antibodies to tick-borne encephalitis virus (TBEV) with NLuc fragments capable of complementation with restoration of bioluminescent activity were obtained. The possibility of detecting TBEV in model samples and tick extracts using single-phase bioluminescence analysis has been demonstrated.

The developed version of NLuc luciferase has properties similar to intact NanoLuc and does not lose bioluminescent activity when chemically conjugated with various biospecific molecules. The conjugates obtained on its basis were tested in molecular microassay of a number of diagnostically important targets (cardiac and tumor markers, as well as corpuscular antigen) and showed good prospects of their use for the development of test systems based on them.

Funding: The study was funded by State Assignment of the Ministry of Science and Higher Education of the Russian Federation (projects No. FWES-2022-0002).

Список литературы/References

1. Krasitskaya V.V., Bashmakova E.E., Frank L.A. Coelenterazine-dependent luciferases as a powerful analytical tool for research and biomedical applications. *Int J Mol Sci.* 2020;21(20):7465
2. Kudryavtsev A.N., Krasitskaya V.V., Efremov M.K. et al. Ca^{2+}-Triggered coelenterazine-binding protein *Renilla*: expected and unexpected features. *Int J Mol Sci.* 2023;24(3):2144. doi 10.3390/ijms2403214. 4
3. Krasitskaya V.V., Efremov M.K., Frank L.A. Luciferase NLuc site-specific conjugation to generate reporters for in vitro assays. *Bioconjug Chem*. 2023;34(7):1282-1289. doi 10.1021/acs.bioconjchem.3c00165

doi 10.18699/bgrs2024-9.2-18

Колориметрическая детекция белка DKK-1 с использованием ДНК-аптамеров

Шатунова Е.А.[1]*, Королев М.А.[1, 2], Воробьева М.А.[1]

[1] *Институт химической биологии и медицины СО РАН, Новосибирск, Россия*

[2] *Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия*

* *lizashatunova@yandex.ru*

Ключевые слова: ДНК-аптамеры; Dickkoppf-1; колориметрический анализ; анкилозирующий спондилит

Мотивация и цель: Сывороточный белок Dickkopf-1 (DKK-1) является одним из ингибиторов Wnt-сигнального пути и важным регулятором костного метаболизма. В частности, его рассматривают в качестве потенциального сывороточного маркера при диагностике иммуновоспалительных ревматических заболеваний, в том числе анкилозирующего спондилита. Соответственно, существует потребность в тест-системах для определения DKK-1 в сыворотке, позволяющих проводить регулярный мониторинг уровня этого белка и сохраняющих воспроизводимость свойств в ходе долговременных исследований. Перспективным подходом является создание ИФА-подобных систем на основе аптамеров – синтетических одноцепочечных молекул РНК и ДНК, способных селективно и с высокой аффинностью связывать специфичные мишени.

Методы и алгоритмы: Для получения аптамеров использовали in vitro селекцию из комбинаторной библиотеки оцДНК с неравномерной рандомизацией случайной области с последующим высокопроизводительным секвенированием на платформе MiSeq Illumina в ЦКП «Геномика» ИХБФМ СО РАН. Для химического синтеза комбинаторной оцДНК-библиотеки, праймеров и аптамеров использовали твердофазный фосфитамидный метод. Скрининг аффинности кандидатных аптамеров проводили с использованием колориметрического метода в микропланшетном варианте с нековалентной иммобилизацией белка на поверхность лунки или биослойной интерферометрии. При конструировании колориметрических тест-систем для определения белка DKK-1 аптамер с концевой алифатической аминогруппой ковалентно иммобилизовали на карбокси-активированную поверхность планшета, после внесения белка добавляли конъюгат антитела с пероксидазой или биотинилированный аптамер и конъюгат стрептавидин-пероксидаза, для получения сигнала использовали хромогенный субстрат TMB.

Результаты: После четырех раундов отбора и секвенирования пять наиболее представленных в финальной библиотеке аптамеров были выбраны для синтеза и скрининга аффинности. Показано, что 71-звенные ДНК-аптамеры DK1, DK2, DK3, DK4 связывают белок DKK-1 с высокой аффинностью (константы диссоциации в диапазоне 1.3–3.7 нМ). Создана серия укороченных аптамеров (41–50 нт), сохранивших высокое сродство к DKK-1. Аптамеры DK1, DK4, укороченные варианты DK1_48t и DK4_41t и ДНК-аптамер TD10, полученный на основе литературных данных [1], использовали для конструирования

колориметрических систем детекции DKK-1 в микропланшетном варианте. Показано, что в модельных условиях для тест-систем сэндвич-формата «аптамер/антитело» в диапазоне 0.156–10 нг/мл DKK-1 при увеличении концентрации белка наблюдается стабильный рост сигнала, максимальную амплитуду сигнала обеспечивали аптамеры DK4 и TD10. Исследована возможность создания колориметрических тест-систем сэндвич-формата «аптамер/аптамер» без использования антител к белку. Выявлены пары аптамеров, подходящие для систем такого типа, и продемонстрирована их работоспособность в модельном растворе и в допированной DKK-1 сыворотке крови в диапазоне концентраций 0.156-10 нМ. При этом наилучшими характеристиками обладали системы с использованием пар аптамеров DK1/DK4 и DK4_41t/TD10. Дальнейшие исследования будут направлены на дополнительную оптимизацию колориметрических систем для повышения чувствительности в образцах сыворотки.
Выводы: Получена серия новых высокоаффинных ДНК-аптамеров к DKK-1 и их укороченных вариантов, сохраняющих сродство к DKK-1. На основе сэндвич-пар ДНК-аптамеров были сконструированы ИФА-подобные системы колориметрической детекции DKK-1, показана их принципиальная работоспособность.
Финансирование: Исследование поддержано грантом РНФ и правительства НСО № 22-15-20050.

Colorimetric detection of DKK-1 protein using DNA aptamers

Shatunova E.[1]*, Korolev M.[1, 2], Vorobyeva M.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *lizashatunova@yandex.ru*

Key words: DNA-aptamers; Dickkopf-1; colorimetric detection; ankylosing spondylitis

Motivation and Aim: The serum protein Dickkopf-1 (DKK-1) is a Wnt-signaling pathway inhibitor and a key regulator of bone metabolism. Particularly, it is considered a potential serum marker for the diagnosis of immunoinflammatory rheumatic diseases, such as ankylosing spondylitis. Therefore, there is a need for test systems for measuring DKK-1 in serum that support regular monitoring of this protein's concentration and maintain reproducible properties during long-term studies. The development of ELISA-like systems based on aptamers, synthetic single-stranded RNA and DNA molecules with high affinity for selective binding of certain targets, is a promising strategy.
Methods and Algorithms: DNA aptamers were in vitro selected from a combinatorial ssDNA library with uneven randomization. The selection was followed by high-throughput sequencing on the MiSeq Illumina platform at the Genomics Core Facility of the ICBFM SB RAS. The solid-phase phosphoramidite method was used for the chemical synthesis of a combinatorial ssDNA library, primers, and aptamers. The affinity screening of candidate aptamers was carried out using a colorimetric microplate method with non-covalent immobilization of the protein on the well surface or biolayer interferometry. To construct a colorimetric test system for measuring the DKK-1 protein,

an aptamer with a terminal aliphatic amino group was covalently immobilized on the carboxy-activated surface of the plate. After adding the protein, an antibody-peroxidase conjugate or a biotinylated aptamer and a streptavidin-peroxidase conjugate were added, and the chromogenic substrate TMB was used to obtain a signal.

Results: After four rounds of selection and sequencing, the five aptamers most prevalent in the final library were chosen for synthesis and affinity screening following. 71-mer DNA aptamers DK1, DK2, DK3, and DK4 bind the DKK-1 protein with high affinity (dissociation constants in the range of 1.3–3.7 nM). We also generated a series of truncated aptamers (41–50 nt) which retained high affinity for DKK-1. Aptamers DK1, DK4, truncated versions DK1_48t and DK4_41t, and previously published DNA aptamer TD10 [1], were used for the construction of colorimetric DKK-1 detection systems in a microplate format. Under model conditions, the aptamer/antibody variant of sandwich test systems provided a stable signal increase in the signal with the rise of protein concentration in the range of 0.156–10 ng/ml DKK-1; the maximum signal amplitude was obtained for the DK4 and TD10 aptamers. We also investigated the possibility to use the aptamer/aptamer sandwich format of the DKK-1 colorimetric detection without DKK-1-specific antibodies. We identified pairs of aptamers suitable for systems of this type, and their demonstrated their applicability in both a model solution and in serum spiked with 0.156–10 nM of DKK-1. The best characteristics were provided by systems made of DK1/DK4 and DK4_41t/TD10 aptamer sandwich pairs. Future research will focus on further optimization of colorimetric systems to improve their sensitivity in serum samples.

Conclusion: We generated new high-affinity DNA aptamers for DKK-1 and their truncated variants that retain affinity for DKK-1. ELISA-like colorimetric detection systems for DKK-1 based on sandwich pairs of DNA aptamers were constructed, and their principal functionality was demonstrated.

Funding: The study is supported by the grant of Russian Science Foundation and government of Novosibirsk region No. 22-15-20050.

Список литературы/References

1. Zhou Y. et al. Developing slow-off dickkopf-1 aptamers for early-diagnosis of hepatocellular carcinoma. *Talanta*. 2019;194:422-429

doi 10.18699/bgrs2024-9.2-19

Big data-based studies on neurodegenerative-related interactome, aging and longevity

Chen M.
Zhejiang University, Zhejiang, China

Key words: neurodegenerative disease; protein–protein interaction; biological age; aging; longevity; machine learning; bioinformatics

Abstract: Latest trends in bioinformatics and computational biology play a crucial role in analyzing large and complex biological datasets, understanding molecular mechanisms of life and disease, and accelerating the development of novel therapies. This talk focuses on the neurodegenerative-related interactome, machine learning-based biological age prediction, and human aging and longevity knowledge graph.
We conducted a comprehensive analysis of neurodegenerative disease-related proteins and their interactions, generating a high-resolution network with structural information. The Neurodegenerative Disease Atlas (NDAtlas) was developed, allowing for 3D molecular graphics beyond traditional 2D network information. We proposed a composite machine learning-based biological age (ML-BA) model based on biomarkers obtained from medical examination data. The composite ML-BA is strongly associated with healthy risk indicators and various diseases, providing improved aging measurement capabilities and supporting the potential application of machine learning in aging research. We introduced HALD, a text mining-based human aging and longevity knowledge graph containing essential entities in the field of aging and longevity and related literature curated from PubMed. HALD enables a comprehensive understanding of aging and longevity mechanisms, providing a foundation for developing anti-aging therapies for aging-related diseases.

References

1. Chen H., Zhou Y., Liu Y., Zhang P., Chen M. Network integration and protein structural binding analysis of neurodegeneration-related interactome. *Brief Bioinform.* 2023;24(4):bbad237
2. Wu Z., Feng C., Hu Y. et al. HALD, a human aging and longevity knowledge graph for precision gerontology and geroscience analyses. *Sci Data*. 2023;10(1):851
3. Hu X., Feng C., Zhou Y., Harrison A., Chen M. DeepTrio: a ternary prediction system for protein-protein interaction using mask multiple parallel convolutional neural networks. *Bioinformatics*. 2022;38(3):694-702
4. Yang Q., Gao S., Lin J. et al. A machine learning-based data mining in medical examination data: a biological features-based biological age prediction model. *BMC Bioinformatics*. 2022;23:411
5. Zhou Y., Chen H., Li S., Chen M. mPPI: A PPI Database Extension for Visualizing Structural Interactome in One-to-Many Manner. *Database*. 2021;2021:baab036

doi 10.18699/bgrs2024-9.2-20

The genes and eQTLs to unravel the genetic underpinnings of skin wrinkling and sagging

Khvorykh G.[1], Ageeva E.[2]*

[1] *National Research Centre "Kurchatov Institute", Moscow, Russia*

[2] *V.I. Vernadsky Crimean Federal University, Simferopol, Republic of Crimea, Russia*

* *ageevaeliz@rambler.ru*

Key words: skin aging; single nucleotide polymorphism; genes

Motivation and Aim: Skin aging stands as one of the most evident markers of the aging process, reflecting a complex interplay of genetic predispositions and environmental influences. Unraveling the precise genetic contributions to this phenomenon poses a significant challenge. A study examining expression quantitative trait loci (eQTL) in blood samples from 244 individuals of European descent revealed a profound impact of altered metabolism on longevity. Specially, approximately one-fifth of differentially expressed genes were found to be influenced by genetic variants in cis [1]. In parallel, the results of 44 genome-wide association studies (GWAS) identified 19 single nucleotide polymorphisms (SNPs) significantly associated with skin aging [2]. Notably, these associations were replicated across multiple studies. However, there remains a gap in our understanding regarding how these polymorphisms affect gene function, hindering their translation into clinical applications [3]. Therefore, the primary objective of this research is to elucidate the candidate genes implicated in skin aging, along with their associated SNPs and eQTLs. Such insights are crucial for guiding future associative studies and potentially paving the way for targeted interventions in the field of skin aging research. Skin aging phenotypes are numerous. Recent systematic review listed 56 of them [2]. Besides, ethnic populations demonstrate differences in skin aging phenotypes and genetics [4]. To be specific, in this research we focused on genes and eQTLs of skin wrinkling and sagging in people of European ancestry.

Methods and Algorithms: The genes associated with skin wrinkling and sagging were derived from GWAS published previously and summarized in [2]. We were interested in protein coding genes uniquely presented in the phenotypes of category A. The genomic coordinates of these genes under human genome assembly GRCh37 were retrieved from Ensembl (grch37.ensembl.org, accessed on April 1, 2024) via the web-interface. The cis-eQTL were obtained from Adult Genotype-Tissue Expression (GTEx) v.8 (gtexportal.org, accessed March 26, 2024) for the tissues denoted as Skin_Not_Sun_Exposed_Suprapubic and Skin_Sun_Exposed_Lower_leg. The genome assembly of eQTLs was GRCh38. The eQTLs were annotated using the dbSNP 155 All by submitting genomic coordinates in bed file format to the database via the web interface (genome.ucsc.edu, accessed on April 4, 2024). The start and end of the region in the bed file were defined as x-1 and x, respectively, where x represented the coordinate of the polymorphism in the GTEx data. The obtained data were filtered by polymorphism type, retaining only SNVs. The annotation of eQTLs using the GWAS Catalogue database (https://www.ebi.ac.uk/gwas, accessed on April 11, 2024) was performed applying the API of this resource. Annotation of eQTLs using the Clinvar database was carried out applying the Entrez Direct utility set (accessed on April 12, 2024). The lists

of diseases associated with the genes of interest, eQTLs, and SNPs were obtained by submitting a request via the web form of the disgenet.org portal (accessed on April 14, 2024). Linkage disequilibrium (LD) assessment followed by identification of Gabriel's blocks and tagging SNPs (tagSNPs) was performed using Haploview 4.2 software [6] for a set of genotypic data from 95 unrelated individuals of the CEU population of the 1000 Genomes Project (Phase III) [5]. The genotypic data of this project were downloaded by the link https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502 (accessed on September 16, 2021). The individual genotypes were selected using vcftools 0.17 [7]. Data processing was automated by custom scripts in Python 3.10 [8], R 4.2.1 [9], Perl v5.34.0 [10], and GNU Bash 5.1.16 [11].

Results: The exploratory analysis of GWAS of skin aging resulted in six protein coding genes (MACROH2A2, DLGAP1, NEDD9, MYH11, OLFM1 and VAV3) uniquely associated with wrinkling and sagging and supported by whole genome level of significance (p-value < 5×10^{-08}). Four of these genes contained eQTLs (MACROH2A2 – 81, DLGAP1 – 1, OLFM1 – 21, and VAV3 – 53). The genes OLFM1 (Fig. 1) and VAV3 reveled quite monolithic blocks of LD resulting in 4 and 3 tagSNPs. The gene MACROH2A2 showed the highest variability of LD resulting in 11 Gabriel's blocks and 34 tagSNP. Some of eQTL being in DisGeNET and Clinvar databases can be taken into account while choosing tagSNP for subsequent analysis.

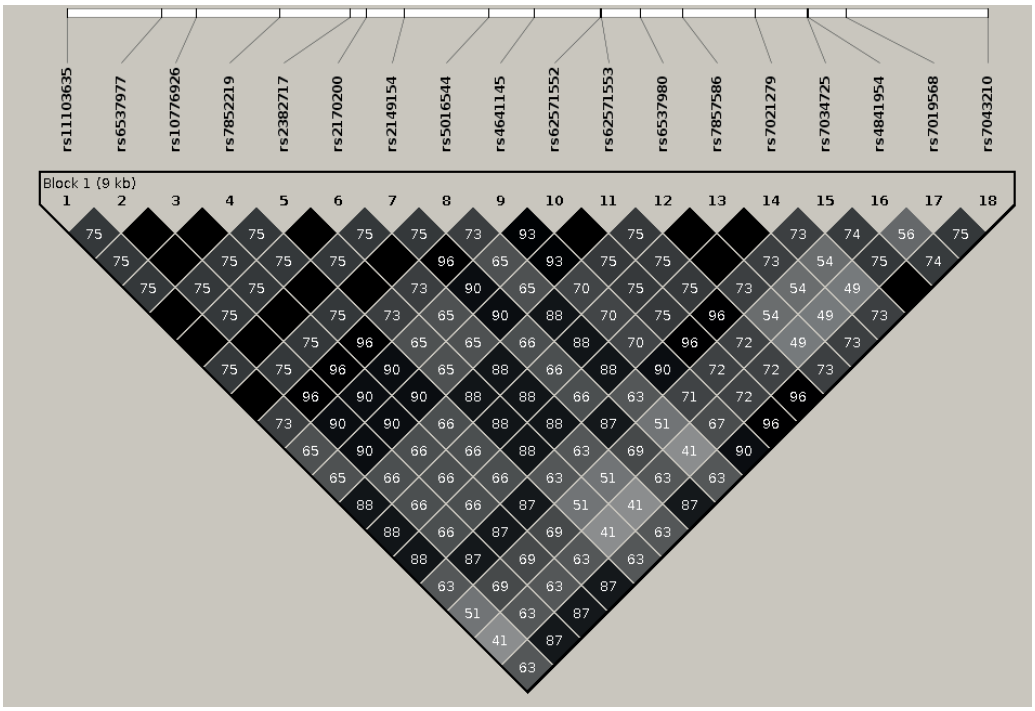


Fig. 1. The heatmap of LD (r2 values) for eQTLs of OLFM1

Conclusion: Using open data, we identified 6 candidate genes of skin wrinkling and sagging, characterized by eQTLs and possible tagSNPs. The results of the analysis made contribute to the transfer of GWAS signals into clinical settings.

References

1. Häsler R., Venkatesh G., Tan Q. et al. Genetic interplay between human longevity and metabolic pathways – a large-scale eQTL study. *Aging Cell*. 2017;16:716-25
2. Ng J.Y., Chew F.T. A systematic review of skin ageing genes: gene pleiotropy and genes on the chromosomal band 16q24.3 may drive skin ageing. *Sci Rep*. 2022;12:13099. doi 10.1038/s41598-022-17443-1
3. Li B., Ritchie M.D. From GWAS to Gene: Transcriptome-Wide Association Studies and Other Methods to Functionally Understand GWAS Discoveries. *Front Genet*. 2021;12:713230
4. Makrantonaki E., Bekou V., Zouboulis C.C. Genetics and skin aging. *Dermatoendocrinol*. 2012;4:280-4. doi 10.4161/derm.22372
5. Auton A., Abecasis G.R., Altshuler D.M. et al. A global reference for human genetic variation. *Nature*. 2015;526:68-74. doi 10.1038/nature15393
6. Barrett J.C., Fry B., Maller J., Daly M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21:263-265. doi 10.1093/bioinformatics/bth457
7. Danecek P., Auton A., Abecasis G. et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156-2158. doi 10.1093/bioinformatics/btr330
8. Van Rossum G., Drake F.L. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace, 2009
9. R Core Team. R: A Language and Environment for Statistical Computing, 2021. https://www.r-project.org/
10. Wall L., Christiansen T., Orwant J. Programming perl. O'Reilly Media, Inc., 2000
11. GNU P. Free Software Foundation. Bash (5.1) [Unix shell program]. 2020. https://www.gnu.org/software/bash/

doi 10.18699/bgrs2024-9.2-21

Proatherogenic potential of oxidative modifications in LDL

Kiseleva D.[1, 2]*, Ziganshin R.[3], Kuzmin V.[4], Cherednichenko V.[1], Khovantseva U.[1], Bogatyreva A.[1], Markina Yu.[1], Markin A.[1, 5]

[1] *Laboratory of Cellular and Molecular Pathology of Cardiovascular System, Petrovsky National Research Centre of Surgery, Moscow, Russia*

[2] *Department of Biophysics, Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia*

[3] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia*

[4] *Biotechnological Faculty, Lomonosov Moscow State University, Moscow, Russia*

[5] *Medical Institute, Peoples' Friendship University of Russia Named after Patrice Lumumba (RUDN University), Moscow, Russia*

* *kiseleva.dg@gmail.com*

Key words: LDL oxidation; protein oxidation; post-translational modifications; atherosclerosis; cardiovascular disease; foam cells; inflammation

Motivation and Aim: Atherosclerosis is a chronic vascular disease and is one of the main causes of acute cardiovascular events. All currently existing theories that explain the process of the initial atherosclerotic lesion formation underline the important role of low-density lipoprotein (LDL) in the development of this pathology. For more than 20 years, it has been emphasized that not LDL itself, but rather its oxidized modifications (oxLDL) [1], cause an increased accumulation of cholesterol in the cells of the subendothelial intimal layer, including macrophages, which occurs through specialized LOX-1 receptors [2]. However, it is difficult to conduct experiments with oxLDL *in vitro* for the following reasons:

- Actual oxLDL level represent a small percentage of total LDL in plasma;
- In order to obtain a sufficient amount of oxLDL for research purposes, the native LDL fraction from the blood plasma of healthy donors is artificially modified, for example, by $CuSO_4$ or lipoxygenase, which has a destructive effect on both lipids and proteins in LDL, and cannot reproduce actual modifications found in patients;
- Experimental data comparing the effect of native, aggregated and oxLDL show conflicting results regarding the accumulation of cholesterol in cells, without confirming the existing hypothesis;

The measurement of oxLDL in the plasma of patients and the correlations with clinical data, namely the risk for development of atherosclerosis is based on enzyme-linked immunosorbent assay (ELISA), i.e. on protein modifications. Each ELISA kit has various antigens, moreover, the modification of only Apolipoprotein B (ApoB) is most frequently measured.

It is also known that patients with type 2 diabetes mellitus (T2DM) suffer from an active progression of atherosclerosis, which increases mortality from cardiovascular diseases (CVD) by 4–5 times in patients in this group. Hyperglycemia promotes the production of reactive oxygen species in mitochondria, increasing the intracellular accumulation of advanced glycation end products, and eventually leads to oxidative modification of LDL [3].

Thus, the purpose of this study was to analyze the existing modifications in native LDL samples isolated from the plasma of patients with T2DM and chronic CVD, and their connection with the accumulation of cholesterol by macrophages *in vitro*.

Methods and Algorithms: LDL were isolated from the plasma of 15 patients using a standard protocol for sequential ultracentrifugation with solutions of various densities (LEC of the Petrovsky National Research Center of Surgery, Approval No. 5, 11 December 2022). The cholesterol-to-protein ratio in THP-1 cells was carried out using the CHOLESTEROL liquicolor kit, the protein level was determined by the Lowry method. Cells were previously stimulated to differentiate into M0 macrophages by phorbol-12-myristate 13-acetate (PMA) treatment. Liquid chromatography and mass spectrometry was performed with an Ultimate 3000 Nano LC System coupled to the Orbitrap Tribrid Lumos mass spectrometer. Data analysis was performed in MaxQuant 2.4.2.0, Perseus 2.0.11 and with Python.

Results: 38 post-translational modifications as well as their combinations in 151 proteins were analyzed. The results are presented in Fig. 1.

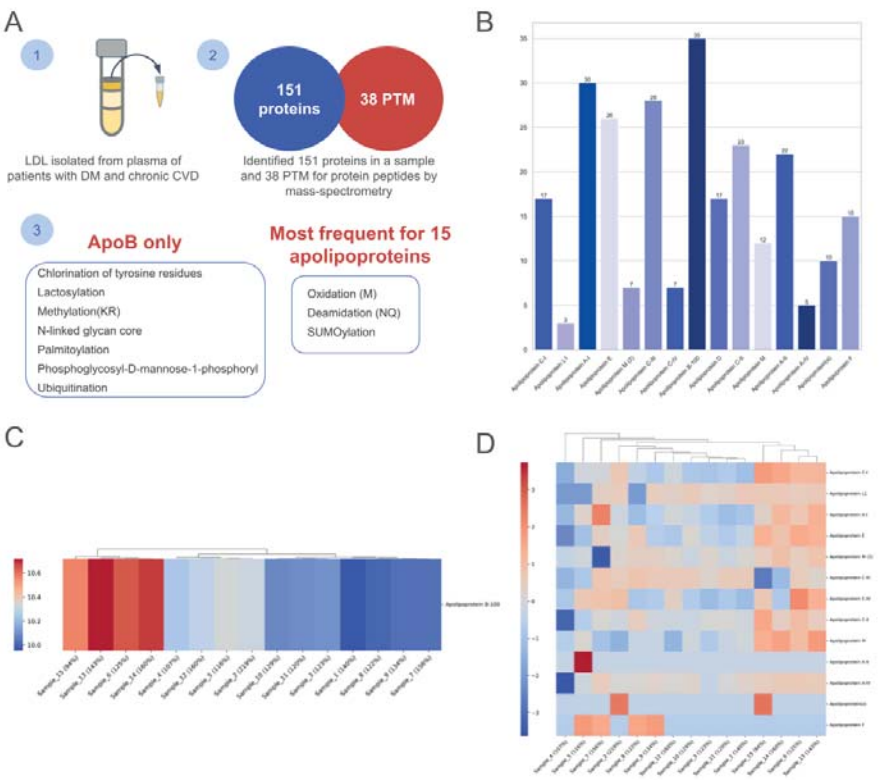


Fig. 1. Post-translational modifications (PTM) of 15 apolipoproteins detected in LDL samples from patients with T2DM and chronic CVD. A. General approach to the analysis; 38 modifications were found in 151 proteins. The figure shows the most common modifications in 15 apolipoproteins, and specifies those that are unique to ApoB. B. Barplot illustrates the frequency of PTM in apolipoproteins, i.e. the number of modification types that occur in the protein. C. Clustermap for the oxidation of Met, as the most common modification of ApoB in LDL samples depending on the accumulation of cholesterol by cells. D. Clustermap for the oxidation of Met, as the most common modification of apolipoproteins (excepting ApoB) in LDL samples depending on the accumulation of cholesterol by cells

LDL particles have a complex morphology. ApoB is the main structural protein for the lipid spherical particle, however, other apolipoproteins, as well as certain blood plasma proteins, also bind to the surface of LDL. Since the apolipoprotein group is the major group of LDL structure, this work shows results for 15 of the 151 proteins detected. The main modifications that occur in these proteins are shown in Fig. 1A. Apolipoproteins B, A-I, C-III and E are most frequently susceptible to modification (Fig. 1B). The most common modification found among all apolipoproteins is the oxidation of Met, rather than Cys or Lys, which are detected after *in vitro* artificial modification of LDL or measured by ELISA kits [4]. However, neither ApoB nor the other apolipoproteins showed a relationship between the percentage of cholesterol accumulation in cells and the quantity of Met residues oxidation (Fig. 1C, D). Other modifications that are associated with protein oxidation [5, 6] also do not affect the percentage of cholesterol accumulation by macrophages. Statistically significant Pearson correlations were found between the percentage of accumulation and the relative abundance (rIBAQ) of such proteins in the sample as DAP kinase 1 (0.74), Myomegalin (0.71), Suprabasin (0.67) and Spearman correlation for Fibronectin (0.68).

Conclusion: Out of the 38 detected post-translational modifications in LDL that could occur during the oxidation of methionine, lysine, cysteine, tyrosine or other amino acids, none of them affected the increased accumulation of cholesterol by macrophages. The authors suggest that accumulation and inflammation are induced by other proteins that form an LDL particle, which may include: DAP kinase 1, Myomegalin, Suprabasin, Fibronectin.

Funding: This study is supported by the Ministry of Science and Higher Education of the Russian Federation (No. 123030700024-4).

References

1. Steinberg D., Parthasarathy S., Carew T.E., Khoo J.C., Witztum J.L. Beyond cholesterol. Modifications of low-density lipoprotein that increase its atherogenicity [see comments]. *N Engl J Med*. 1989;320:915-924
2. Akhmedov A., Sawamura T., Chen C.H., Kraler S., Vdovenko D., Lüscher T.F. Lectin-like oxidized low-density lipoprotein receptor-1 (LOX-1): a crucial driver of atherosclerotic cardiovascular disease. *Eur Heart J*. 2021;42(18):1797-1807
3. Mao L., Yin R., Yang L., Zhao D. Role of advanced glycation end products on vascular smooth muscle cells under diabetic atherosclerosis. *Front Endocrinol*. 2022;13:983723
4. Yang C.Y., Gu Z.W., Yang M. et al. Selective modification of apoB-100 in the oxidation of low density lipoproteins by myeloperoxidase in vitro. *J Lipid Res*. 1999;40(4):686-98
5. Parthasarathy S., Raghavamenon A., Garelnabi M.O., Santanam N. Oxidized low-density lipoprotein. *Methods Mol Biol*. 2010;610:403-17
6. Ramazi S., Zahiri J. Posttranslational modifications in proteins: resources, tools and prediction methods. *Database (Oxford)*. 2021;2021:baab012

doi 10.18699/bgrs2024-9.2-22

Effect of hypertrophic mutations of cardiac myosin-binding protein C on actin-myosin interaction

Kochurova A.[1], Beldiia E.[1, 2], Matyushenko A.[3], Bershitsky S.[1], Kopylova G.[1]*, Shchepkin D.[1]

[1] *Institute of Immunology and Physiology, UB RAS, Yekaterinburg, Russia*

[2] *UrFU, Ekaterinburg, Russia*

[3] *Federal Research Center of Biotechnology, RAS, Moscow, Russia*

* *g_rodionova@mail.ru*

Key words: cardiac myosin binding protein C; hypertrophic mutations; myocardium; omecamtiv mecarbil; calcium regulation; actin-myosin interaction; *in vitro* motility assay

Motivation and Aim: About 45 % of sarcomere protein mutations associated with hypertrophic cardiomyopathies (HCM) are cardiac myosin binding protein C (cMyBP-C). The molecular mechanisms of the development of HCM with point mutations in cMyBP-C are practically not studied. We studied the effect of HCM mutations D75N, P161S, and L352P in the N-terminal part of the cMyBP-C on the actin-myosin interaction in the myocardium.

Methods: Myosin was obtained from the left ventricle of the pig heart. C0-C2 fragment of wild-type (WT) human cMyBP-C and with HCM mutations D75N, P161S, and L352P, human cardiac tropomyosin and troponin were expressed in *E. coli*.

The effect of mutations on actin-myosin interaction was studied in an *in vitro* motility assay. The calcium dependence of the sliding velocity of thin filaments, reconstructed from F-actin, tropomyosin, and troponin, was fitted by the Hill equation: $V = V_{max} \times (1+10^{n(pCa-pCa50)})^{-1}$, where V and V_{max} are a velocity and the maximum velocity at saturating calcium concentration respectively; $p\mathrm{Ca}_{50}$ (i.e., calcium sensitivity) is pCa at which half-maximal velocity is achieved; and n is the Hill cooperativity coefficient.

Results: We found that 500 nM P161S C0-C2 fragment increased the maximum sliding velocity of thin filaments at a saturating Ca^{2+} concentration by 30 %, and mutations D75N and L352P reduced it by 20 and 70 % compared to the WT C0-C2 fragment. The P161S and D75N mutations significantly reduced the Ca^{2+} sensitivity of the actin-myosin interaction. At 500 nM L352P C0-C2, the filament velocity did not depend on the Ca^{2+} concentration. At 300 nM L352P C0-C2, this dependence was sigmoidal, its $p\mathrm{Ca}_{50}$ was lower than with 300 nM WT C0-C2; and mutation L352P did not inhibit the filament sliding at low Ca^{2+} concentrations. The P161S and D75N mutations impaired thin filament activation by increasing the myosin concentration at which filament velocity is half-maximal.

Then we studied the effect of a specific activator of ventricular myosin, omecamtiv mecarbil (OM) [1, 2], on actin-myosin interaction in the presence of the P161S cMyBP-C mutation. Previously, it was shown that OM increases the force of multicellular preparations from the left ventricle at non-saturating calcium concentration [3, 4]. We hypothesized that OM could neutralize the decrease in the Ca^{2+} sensitivity of actin-myosin interaction. We found that in the absence of C0-C2 fragment, OM halved the maximum sliding velocity of thin filaments and increased the Ca^{2+} sensitivity by 0.2 pCa. With WT C0-C2 fragment, OM reduced maximum filament velocity to a lesser

extent, only by 30 %, and significantly increased Ca^{2+} sensitivity. With P161S C0-C2 fragment, OM decreased the filament velocity by 30 %, but calcium sensitivity did not change.

Conclusion: The studied HCM cMyBP-C mutations disrupt the calcium regulation of actin-myosin interaction, which may be one of the molecular mechanisms of HCM development. The P161S mutation impairs the activation of thin filaments, and OM only partially neutralizes the negative effect of this mutation on the actin-myosin interaction.

Funding: The study is supported by the Russian Science Foundation (No. 22-14-00174).

References

1. Teerlink J.R. A novel approach to improve cardiac performance: cardiac myosin activators. *Heart Fail Rev*. 2009;14(4):289-298. doi 10.1007/s10741-009-9135-0
2. Felker G.M., Solomon S.D., Claggett B. et al. Assessment of omecamtiv mecarbil for the treatment of patients with severe heart failure: a post hoc analysis of data from the galactic-hf randomized clinical trial. *JAMA Cardiol*. 2022;7(1):26-34. doi 10.1001/jamacardio.2021.4027
3. Nagy L., Kovács Á., Bódi B. et al. The novel cardiac myosin activator omecamtiv mecarbil increases the calcium sensitivity of force production in isolated cardiomyocytes and skeletal muscle fibres of the rat. *Br J Pharmacol*. 2015;172(18):4506-4518. doi 10.1111/bph.13235
4. Mamidi R., Li J., Gresham K.S. et al. Dose-dependent effects of the myosin activator omecamtiv mecarbil on cross-bridge behavior and force generation in failing human myocardium. *Circ Heart Fail*. 2017;10(10):e004257. doi 10.1161/CIRCHEARTFAILURE.117.004257

doi 10.18699/bgrs2024-9.2-23

Biological consequences of cyanobacteria and microalgae

Lykov A.[1]*, Gevorgiz R.[2], Zheleznova S.[2], Uvarov I.[3], Poveshchenko O.[1]

[1] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *A.O. Kovalevsky Institute of Biology of the Southern Seas, RAS, Sevastopol, Russia*

[3] *Koltsovskaya Veterinary Clinic, Koltsovo, Russia*

* *aplykov2@mail.ru*

Key words: microalgae; cyanobacteria; Fucoxanthin; cytotoxicity; antimicrobial activity

Motivation and Aim: Microalgae and cyanobacteria are a diverse group of microorganisms that live in a variety of settings, are capable of oxygenic photosynthesis, and can create a thick microbial mat even in harsh conditions. Microalgae and cyanobacteria contain a diverse variety of physiologically active compounds that can be exploited to create promising dietary supplements and anti-inflammatory medications [1–3]. Only a few of the more than 12,000 microalgae and 5,000 cyanobacteria species are approved for human consumption. On this basis, it appeared natural to investigate the effects of numerous microalgae and cyanobacteria bioactive substances *in vitro* and *in vivo*.

Methods and Algorithms: Green microalgae (*Chlorella vulgaris* Beijerink, *Coelastrella sp*., and *Tetraselmis (Platymonas) viridis* Rouchijajnen), cyanobacteria (*Arthrospira* (Spirulina) *platensis* (Nordstedt) Gomont, *Leptolyngbya cf. ectocarpi*, *Planktothrix agardhii*, and *Roholtiella mixta sp. nov.),* diatoms (*Cylindrotheca closterium* (Ehrenb.) Riemann et Lewin, and *Nanofrustulum shiloi*) and reds microalgae (*Porphyridium purpureum* (Bory) Drewet Ross) were obtained from the Kovalevsky Institute of Biology of the Southern Seas culture collection. Sample extracts were made from dry biomass by passive diffusion of physiologically active chemicals into a 1 % dimethyl sulfoxide (DMSO) solution for 24 hours or into vegetable oil for 72 hours at 37 °C. The effect of oil extracts of microalgae and cyanobacteria was assessed in an experiment on mice fed conventional pelleted food pre-soaked in oil extract; the animals were withdrawn from the experiment on the 13th day. Blood serum was taken, lymphocytes were separated from the thymus and spleen, and cell functional characteristics were assessed in vitro. An extract from kidney and liver homogenate was also obtained. The levels of IL-1β, IL-10, and TNFα were measured in serum, conditioned medium, and kidney and liver tissue extracts using commercial kits "Interleukin-1 beta-IFA-BEST", "Interleukin-10-IFA-BEST", and "TNF-IFA-BEST alpha" (Vector-Best, Russia), according to the manufacturer's instructions. The quantities of persistent nitric oxide metabolites in serum, conditioned medium, and kidney and liver tissue extracts were measured using Griess reagent. The cytotoxicity of DMSO extracts of materials was assessed against mouse splenocytes and peritoneal macrophages, as well as human and rat bone marrow mesenchymal stem cells. Furthermore, the antimicrobial potential of DMSO extracts of samples against Gram-positive and Gram-negative strains of bacteria, as well as *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* was assessed using an *in vitro* micro method, and the *in vivo* effect of an alcoholic extract of Fucoxanthin in an experimental tuberculosis model in mice was evaluated. Furthermore,

the impact of intragastric administration of alcoholic and oil Fucoxanthin on blood biochemical markers and cytokines was investigated. Data is presented in the form of mean and standard deviation (M±SD) in the tables. Utilizing Statistica 10.0 for Windows, data was examined. Statistical significance of differences between samples was evaluated using one-factor analysis of variance (ANOVA) with Bonferroni post hoc test, and accepted at $p < 0.05$.

Results: Microalgae and cyanobacteria can impact the functional characteristics of lymphocytes and liver cells [4]. Oil extracts of microalgae and cyanobacteria have been shown to cause changes in serum levels of albumin (*Chlorella*, Spirulina, *Cylindrotheca*); low-density lipoprotein *(Chlorella*, Spirulina, *Cylindrotheca*); high-density lipoprotein (*Chlorella*, *Cylindrotheca*); total cholesterol (*Chlorella*, Spirulina, *Porphyridium*); progesterone (*Chlorella*, Spirulina, *Cylindrotheca*, *Coelastrella*) and increased albumin levels (*Coelastrella*, *Porphyridium*). In addition, there was a rise in IL-1β (*Chlorella*), TNF-α (*Coelastrella*), IL-10 (*Chlorella*, *Coelastrella*), NO (*Porphyridium*), and decrease in TNF-α levels (*Chlorella*, *Coelastrella*).

In liver tissues, protein (*Coelastrella*, *Porphiridium*), transaminases (*Porphiridium*), IL-1β, and NO (*Porphiridium*) increased while protein (C*ylindrotheca*), IL-1β, TNF-α, and IL-10 decreased (*Chlorella*, Spirulina). The study found increased levels of protein (Spirulina, *Cylindrotheca*, *Porphiridium*), creatinine (C*ylindrotheca*, *Porphiridium*), urea (*Chlorella*, *Cylindrotheca*) in kidney tissues.

A decrease in spontaneous IL-10 synthesis by splenocytes was seen in the *Chlorella*, *Coelastrella*, Spirulina, and *Cylindrotheca* groups, whereas mitogen increased IL-10 production in all experimental groups of animals. *Cylindrotheca* increased IL-1β production by splenocytes, but *Chlorella*, Spirulina, and *Porphiridium* decreased it. Additionally, Spirulina lowered splenocytes' production of TNF-α. Thymocytes in the Spirulina and *Cylindrotheca* group produced more IL-1β, IL-10, and TNFα compared to other experimental groups. Mitogen boosted IL-1β synthesis (*Chlorella*, Spirulina, *Porphiridium*) while decreasing IL-10 production (*Coelastrella*, *Cylindrotheca*) in thymocytes.

Microalgae and cyanobacteria, with their ability to resist and protect against pathogens, are a valuable source of such molecules [5]. A significant cytotoxic activity tested microalgae and cyanobacteria extracts on splenocytes were obtained. DMSO extracts of microalgae and cyanobacteria enhanced the proliferation and NO secretion of human and rat bone marrow mesenchymal stem cells, except *Nanofrustulum*.

Fucoxanthin from Spirulina and *Nanofrustulum* decreased the growth of Gram-positive (*S. aureus, E. faecalis, S. pyogenes*) and Gram-negative (*Kl. pneumoniae, A. baumannii, Ps. aeruginosa*) bacteria compared to control and Ceftazidime on days 14 and 21, especially on day 14. DMSO extracts of microalgae and cyanobacteria, as well as Spirulina Fucoxanthin, suppress the growth of *Kl. pneumoniae*, *S. aureus* MRSA, *A. baumannii*, *E. faecalis* (excluding Fucoxanthin), and *Ps. aeruginosa*, with the exception of *Planktothrix* and *Leptolyngbya*. In comparison to Chlorophyllipt and Rifampicin, DMSO extracts of microalgae and cyanobacteria, particularly *Nanofrustulum*, *Leptolynbya*, and Fucoxanthin, shown antimycotic activity. Furthermore, administering Fucoxanthin alcoholic solution intragastrically to mice with tuberculosis caused by *Mycobacterium tuberculosis* strain H37Rv reduces colony counts in the lungs and spleens.

Fucoxanthin is a carotenoid with fat-burning, anti-inflammatory, antioxidant, anti-infectious, and anti-tumor properties [6]. Intragastric administration of Fucoxanthin

alcohol solution reduces albumin, uric acid, creatinine, cholesterol, high-density lipoprotein, hepatic transaminases, and IL-6, IL-10, and INFγ levels in serum while increasing triglycerides, low-density lipoprotein, and TNFα. In turn, Fucoxanthin oil solution increases serum creatinine, low-density lipoproteins, urea, and TNFα while decreasing albumin, uric acid, triglycerides, cholesterol, high-density lipoproteins, hepatic transaminases, lactate, IL-1β, IL-6, IL-10, and INFγ.

Conclusion: Thus, microalgae and cyanobacteria from various systematic groups influence the levels of proteins, fats, aminotransaminase activity, sex hormones, and filtration function in rodents, as well as the balance of pro-inflammatory and anti-inflammatory cytokines in blood serum, immunocyte conditioned media (mature and naive), kidney, and liver tissues, which should be considered when selecting microalgae and cyanobacteria as an additive in animal and human diets. Data present here showed that extracts from different microalgae and cyanobacteria taxa, especially Fucoxanthin were effective against bacteria (Gram-positive and Gram-negative) and mycobacteria (*Mycobacterium tuberculosis* strain H_{37}Rv and *Mycobacterium smegmatis*). However, it must be emphasized that some microalgae potentiated bacteria growth.

Funding: The work was performed within the framework of the state assignment registration No. 122022800132-1, 1023032700554-2-1.6.16, and 121030300149-0.

References

1. Laroche C. Exopolysaccharides from microalgae and cyanobacteria: diversity of strains, production strategies, and applications. *Mar Drugs*. 2022;20(5):336. doi 10.3390/md20050336
2. Ferrazzano G.F. et al. Cyanobacteria and microalgae as sources of functional foods to improve human general and oral health. *Molecules*. 2020;25(21):5164. doi 10.3390/molecules25215164
3. Alsenani F. et al. Evaluation of microalgae and cyanobacteria as potential sources of antimicrobial compounds. *Saudi Pharm J*. 2020;28(12):1834-1841. doi 10.1016/j.jsps.2020.11.010
4. Tabarzad M. et al. Anti-inflammatory activity of bioactive compounds from microalgae and Cyanobacteria by focusing on the mechanisms of action. *Mol Biol Rep*. 2020;47(8):6193-6205. doi 10.1007/s11033-020-05562-9
5. Guzmán F. et al. Identification of Antimicrobial Peptides from the Microalgae Tetraselmis suecica (Kylin) Butcher and Bactericidal Activity Improvement. *Mar Drugs*. 2019;17(8):453. doi 10.3390/md17080453
6. Takatani N., Taya D., Katsuki A., Beppu F., Yamano Y., Wada A., Miyashita K., Hosokawa M. Identification of paracentrone in fucoxanthin-fed mice and anti-inflammatory effect against lipopolysaccharide-stimulated macrophages and adipocytes. *Mol Nutr Food Res*. 2021;65(2):e2000405. doi 10.1002/mnfr.202000405

doi 10.18699/bgrs2024-9.2-24

Mesenchymal stem cell proliferation and NO production under cyanobacteria and microalgae condition

Lykov A.[1]*, Gevorgiz R.[2], Zheleznova S.[2], Uvarov I.[3], Poveshchenko O.[1]

[1] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *A.O. Kovalevsky Institute of Biology of the Southern Seas, RAS, Sevastopol, Russia*

[3] *Koltsovskaya Veterinary Clinic, Koltsovo, Russia*

* *aplykov2@mail.ru*

Key words: microalgae, cyanobacteria, mesenchymal stem cells, proliferation, NO

Motivation and Aim: Numerous nutrients, such as proteins, fats (fatty acids, polyunsaturated fatty acids, transisomers of fatty acids), carbohydrates, vitamins (B, K, and E), different carotenoids, and colors can be found in microalgae and cyanobacteria [1]. It has been demonstrated that *Chlorella vulgaris* aqueous extract can function in lieu of fetal embryonic serum in the culture of Chinese hamster ovary cells and mesenchymal stem cells [2, 3]. The goal of this research is to determine whether extracts from cyanobacteria and microalgae can affect the way rat and human mesenchymal stem cells function. The purpose of the study is to: 1) assess how extracts from microalgae and cyanobacteria affect the viability and proliferation of human and rat mesenchymal stem cells (MSCs) *in vitro*; and 2) investigate how these extracts affect the secretory activity of these cells in order to produce nitric oxide (NO) *in vitro*.

Methods and Algorithms: Green microalgae (*Tetraselmis (Platymonas) viridis* Rouchijajnen), cyanobacteria (*Leptolyngbya cf. ectocarpi*, *Planktothrix agardhii*, and *Roholtiella mixta sp. nov.),* and diatoms (*Nanofrustulum shiloi*) were obtained from the Kovalevsky Institute of Biology of the Southern Seas culture collection. Sample extracts were made from dry biomass by passive diffusion of physiologically active chemicals into a 1% dimethyl sulfoxide (DMSO) solution for 24 hours at 37°C.

The impact of DMSO extracts of cyanobacteria and microalgae on MSCs from rat and human bone marrow was assessed. Proliferative potential (MTT method) and NO production (Griss reagent) were assessed, and cell membership to true MSCs was confirmed by immunophenotyping and differentiation in the connective tissue direction. Data is presented in the form of mean and standard deviation (M±SD) in the tables. Utilizing Statistica 10.0 for Windows, data was examined. Statistical significance of differences between samples was evaluated using one-factor analysis of variance (ANOVA) with Bonferroni post hoc test, and accepted at $p < 0.05$.

Results: Except *for N. shiloi* (human MSCs) and phycocyanin (rat MSCs), it was observed that adding extracts of microalgae, cyanobacteria, and fucoxanthin to the culture medium promotes the proliferation/viability of human and rat MSCs. Furthermore, it was discovered that, with the exception of *N. shiloi* and *L. cf. ectocarpi*, extracts of microalgae, cyanobacteria, and phycocyanin increase NO generation in human MSCs in a dose-dependent manner. When phycocyanin, microalgae, and cyanobacteria extracts were added to the culture medium, rat MSCs were likewise accountable for enhanced NO generation; however, no obvious dose-dependent effect was observed.

Conclusion: Since the capacity of these microalgae and cyanobacteria species to be examined had not been before, our results of boosting the proliferation and secretory capabilities of human and rat MSCs with DMSO extracts of these microorganisms are unique.

Funding: The work was performed within the framework of the state assignment registration number 122022800132-1, 1023032700554-2-1.6.16, and 121030300149-0.

References

1. Amran R. H., Jamal M. T., Sayegh F., Bowrji S., Satheesh S. A Mini Review on Biotechnological Potentials of Bioactive Compounds and Bioproducts Isolated from Cyanobacteria. *Acta Biol Marisiensis*. 2023;6(2):62-86. doi 10.2478/abmj-2023-0012
2. Ng J.Y., Chua M.L., Zhang C., Hong S., Kumar Y., Gokhale R., Ee P.L.R. *Chlorella vulgaris* Extract as a Serum Replacement That Enhances Mammalian Cell Growth and Protein Expression. *Front Bioeng Biotechnol*. 2020;8:564667. doi 10.3389/fbioe.2020.564667
3. Saberian M., Shahidi Delshad E. The role of *Spirulina platensis* on the proliferation of rat bone marrow-derived mesenchymal stem cells. *Iran J Microbiol*. 2023;15(1):111-120. doi 10.18502/ijm.v15i1.11925

doi 10.18699/bgrs2024-9.2-25

Searching for conservative targets for siRNA-therapeutics in human rotavirus type A genome

Makashov A.[1]*, Brodskaya A.[1, 2]

[1] *Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

[2] *Smorodintsev Research Institute of Influenza, St. Petersburg, Russia*

* *makashov_aa@spbstu.ru*

Key words: rotavirus; siRNA; comparative virology

Motivation and Aim: Developing effective vaccines and treatment for viruses of various strains and subtypes relies on search of the conserved regions of viral genes and proteins. Our project investigated four rotavirus genes: VP4, VP7, NSP1, and NSP4 as potential targets for RNA interference as an antiviral therapy. Finding conserved gene regions for RNA interference could be a universal solution to counteract multiple virus serotypes. The first two genes encode the main rotavirus surface antigens, which play a key role in the interaction of the virus with the infected cell. The other two genes, NSP1 and NSP4, encode nonstructural rotavirus proteins involved in the interaction of virus and immune system, and are critical for self-assembly of new virions as well.

Methods and Algorithms: 9218 viral genomes presented in the NCBI GenBank database as of February 2024 were analyzed. Sequences of human rotavirus group A strains circulating from 1974 to 2022 were included in the analysis.

Multiple alignments of the obtained sequences were performed using the MAFFT tool [1]. We chose the G-INS-i algorithm as the most suitable for sequences with similar lengths and a high percentage of homology. Fasta file with multiple alignment results was converted into a nexus-file of the seqmagick program for further analysis. Bayesian reconstruction of the selected sequences was performed using the MrBayes program [2]. The number of generations was chosen based on a previously performed computational experiment, where after all generations, the standard deviation of cleavage probabilities was below 0.01 (value chosen based on published data). A cladogram with posterior probabilities for each cleavage and a phylogram with average branch lengths were generated and written into a nexus file. The trees were visualized and edited with the FigTree tool [3].

Results: Nonstructural protein genes NSP1 and NSP4 are characterized by greater homogeneity across all sequences. However, we can note the presence of indel polymorphism and small insertions in the reading frame of 6-9 nucleotides for several strains isolated in the United States and Australia in the last decade. Several extended indels have been shown for NSP4, predominantly in South Asia. Strains circulating in India in 2012–2013 are characterized by specific long stable insertions of up to 50 nucleotides, but are not evolutionarily conserved and are not found anywhere else.

Consensus sequences were constructed using the cons program from the EMBOSS package. Further analysis of short conserved regions satisfying multiple miRNA selection requirements was performed using this consensus sequence.

We also performed a comprehensive analysis searching for conserved regions of 20 to 25 nucleotides each with optimal GC composition and certain nucleotide patterns in order to satisfy the rules of selection of effective inhibitory miRNAs [3].

Conclusion: The trees were visualized and edited with the FigTree tool. Most circulating human RVAs are reported to belong to three evolutionary lineages that differ in genome constellations: Wa-like strains, DS-1-like strains and AU-1-like strains [4].
Two potential conserved regions in the NSP1 gene that could be targets for antiviral miRNA inhibition were selected: 260–300 and 675–750 (positions in the consensus nucleotide sequence obtained by alignment); and 3 potential conserved regions for NSP4:160-210, 360-395, 460-490. In the protein-coding region of the VP4 and VP7 genes, it is not possible to find a 20-nucleotide conserved region. The maximum length of the sequence with stable conservatism and optimal GC composition does not exceed 15 nucleotides. Nevertheless, these regions were also identified for further miRNA selection namely 333–352 and 600-650 for VP4 and 548–665 for VP7. Further in these regions were determined target sequences for RNAi. Experimental approbation of RNAi effectiveness is helding on now.
Funding: The study is supported by RSF Project (No. 22-74-10117).

References

1. Katoh K., Misawa K., Kuma K., Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, *Nucleic Acids Res.* 2002;30(14):3059-3066. doi 10.1093/nar/gkf436
2. Ronquist F., Teslenko M., van der Mark P. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539-542. doi 10.1093/sysbio/sys029
3. Rambaut A. FigTree v1.3.1. Institute of Evolutionary Biology. Edinburgh:University of Edinburgh, 2010
4. Matthijnssens J., Ciarlet M., Heiman E. et al. Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. *J Virol*. 2008;82(7):3204-3219. doi 10.1128/JVI.02257-07
5. Laganà A., Veneziano D., Russo F. et al. Computational design of artificial RNA molecules for gene regulation. *Methods Mol Biol*. 2015;1269:393-412. doi 10.1007/978-1-4939-2291-8_25

doi 10.18699/bgrs2024-9.2-26

Shedding light on cell proliferation: cell culture monitoring using light scattering

Naumenko M.*, Moskalensky A.
Laboratory of Optics and Dynamics of Biological Systems, Novosibirsk State University, Novosibirsk, Russia
** m.naumenko@g.nsu.ru*

Key words: cell proliferation; label-free real-time cell analysis

Motivation and Aim: Investigations of cell growth and division enlighten the underlying mechanisms of cellular proliferation, factors that regulate cell cycle progression, and the responses of cells to various stimuli. These studies play a vital role in various scientific fields including cancer research and pharmacology. By now, various techniques and assays to access cell proliferation exist such as cell counting using hemocytometer/automated cell counters, colorimetric assays (MTT), fluorescence (CyQUANT)/ bioluminescence assays (luciferase enzymatic reaction), electric impedance measurements (xCELLigence), etc. However, the development of cost-effective, non-invasive and easy-to-perform methods to trace cell proliferation is of high need and importance. Herein we focus on the development and validation of the device monitoring dynamics of both adherent and suspension cell cultures using laser light scattering.

Methods and Algorithms: In our work we performed numerous modifications of the device developed in our laboratory before[1]. The initial optical scheme of the device incorporates a 850 nm, 3 mW dot laser module and a set of photodiodes BPW34 organized in a row to be compatible with the width of T25 biological cell flask. As the laser beam passes through the flask at the distance of 1 mm above the flask bottom, the light scattered by the cells is detected by photodiodes allowing us to carry out non-invasive suspension cell monitoring.

However, common biological models utilized in the laboratory practice include both adherent and suspension cell lines. Consequently, we decided to extend the optical part of the device with the second laser of 660 nm, 5 mW illuminating the sample flask bottom to access the dynamics of adherent cell layer formation as well. In the pilot experiment, we investigated the process of time-dependent sedimentation of *E. coli* cells in the course of 3 days. Next, we analyzed HEK293 adherent cell layer formation within the time range of 4 days.

To further validate our method, we decided to (i) determine the dynamic range of our device for cells in suspension using different concentrations of a number of suspension cell lines as a model system (ii) use our device to access the proliferation of several adherent cell lines at the different seeding density in parallel (iii) perform a fluorescent staining of the cell samples in (ii) to verify the data obtained with our device.

Inspired by recent advances in fingerprint recognition by means of frustrated total internal reflection, we proceeded with more sophisticated optical design for adherent cell layer tracking. Currently, we are modifying the instrument with the prism embedded in the device body so that it gets in a tight contact with the plastic bottom of the sample flask once the flask is placed to be measured. As the light enters the prism, it gets

redirected and enters the bottom of the cell flask in a total-internal-reflection mode. The "trapped" light can interact with adherent cells resulting in the increased scattered light intensity detected by photodiode sensors.

Results: The results of the experiments with bacterial cell culture sedimentation showed the expected patterns in light-scattering intensity change over the time for both lasers. We observed a decrease of the scattering signal for upper laser beam and an increase of the signal for lower laser beam due to the formation of sedimented cell layer. Subsequently, we saw the same trends for the adherent cell culture model (HEK293 cells). Next, we performed experiments to determine the dynamic range of the instrument for the suspension cell cultures. Also, we tested the device for various adherent cell cultures. The experimental data allow us to estimate the reproducibility and accuracy of the device measurements.

Conclusion: We have developed a brand-new optical device allowing one to monitor cell proliferation for both adherent and suspension cell cultures in a commonly used T25 flask. When compared to other existing techniques, it has several advantages as it allows to perform non-invasive, cost-effective, label-free cell proliferation studies for both adherent and suspension cell populations simultaneously. The device is intuitive to operate and can be easily transported as it is compact and light (0.2 kg). Also, the important figures of merit of the instrument are continuous readout and a high measurement speed (1 s). Currently, device utilization is limited to single sample measurements. Therefore, we are working on the possible solutions to this challenge. The evolving field of cell proliferation studies includes investigations of cell metastatic transformations[2]. From this perspective, it might be advantageous to use an approach based on total internal reflection to differentiate between adherent and lying on the bottom cell populations. We shall therefore continue optimizing the methods described above to hopefully lay a foundation for novel cell research technologies.

Funding: The study was supported by the Ministry of Science and Higher Education of the Russian Federation (project FSUS-2020-0039).

Acknowledgments: We express our deepest gratitude to Petr Laktionov (Laboratory of Epigenetics, NSU) for providing us with the biological samples, for our fruitful scientific discussions of the device development strategies and his word of advice. Likewise, we also thank

Sergey Kulemzin (Institute of Molecular and Cellular Biology SB RAS) for his valuable and constructive input. Furthermore, we thank the entire team of the Laboratory of Optics and Dynamics of Biological Systems for the support, help and encouraging research environment.

References

1. Litunenko D.N., Moskalensky A.E. Wireless monitoring of cell cultures based on light scattering: a novel optical scheme and portable prototype. *J. Biophotonics.* 2024;17(1):e202300234. doi 10.1002/jbio.202300234.
2. Vargas-Accarino E., Herrera-Montávez C., Ramón Y. Cajal S., Aasen T. Spontaneous Cell Detachment and Reattachment in Cancer Cell Lines: An In Vitro Model of Metastasis and Malignancy. *Int J Mol Sci.* 2021;22(9):4929. doi: 10.3390/ijms22094929.

doi 10.18699/bgrs2024-9.2-27

Features of the contractile characteristics of single cardiomyocytes in the myocardial sleeves of the pulmonary veins of guinea pigs

Shchepkin D.*, Butova X., Myachina T., Simonova R., Kochurova A., Khokhlova A., Kopylova G.

Institute of Immunology and Physiology, RAS, Yekaterinburg, Russia
** cmybp@mail.ru*

Key words: myocardial sleeves; pulmonary veins; left atria; single cardiomyocytes; sarcomere shortening; actin-myosin interaction; protein phosphorylation

Motivation and Aim: Myocardial sleeves of the pulmonary veins (PV) extending from the left atrium (LA) are the major source of ectopic activity causing atrial arrhythmias [1]. The structural and electrophysiological characteristics of the PV myocardium differ from those of the LA. The mechanical activity of cardiomyocytes (CM) affects cardiac excitation-contraction coupling [2], however, the mechanical function of PV CM has not been studied yet. We compared the characteristics of the actin-myosin interaction and sarcomere length (SL) dynamics in mechanically non-loaded single CM between the PV and LA in the guinea pig heart.

Methods: All experiments involving animal care and handling were conducted according to Directive 2010/63/EU of the European Parliament and approved by the Animal Care and Use Committee of the Institute of Immunology and Physiology (protocol No. 02/23 from 6 April 2023).

Single CM from the LA and PV were isolated using a combined technique of Langedorff perfusion with intra-tissue injections [3]. The characteristics of CM morphometry (length, width, and surface area) were measured on a picture of resting CM using 40× magnification with the IonOptix system (IonOptix Corporation, Milton, MA, USA) and FIJI ImageJ software (National Institutes of Health, Bethesda, MD, USA). Changes in sarcomere length (SL) during mechanically non-loaded CM contractions were recorded under steady-state conditions after 5 min of CM contractions at 1 Hz using the IonOptix system [3].

Myosin was obtained from the PV and LA tissues of guinea pigs. The sliding velocity of thin filaments over myosin was measured in an *in vitro* motility [3]. Phosphorylation levels of cardiac myosin binding protein-C (cMyBP-C), regulatory light chains of myosin (RLC), troponin T (TnT), and troponin I (TnI) were assessed using gel-electrophoresis with ProQ Diamont and SYPRO Ruby staining.

The comparisons were made using the unpaired two-tailed Student's t-test (parametric analysis) and Mann-Witney U-test (non-parametric analysis). A p-value of <0.05 was considered to indicate a significant difference between groups. Data are expressed as median and interquartile range.

Results: First, we compared LA vs. PV morphometry and contractility at the single-cell level. We found that PV CM had 1.2-fold longer length ($p = 0.0026$) and smaller width ($p = 0.0048$) than LA CM (unpaired Student's t-test), and the cell surface area was not different between the LA and PV CM ($p = 0.6868$).

Representative traces of sarcomere shortening-relengthening in contracting LA and PV CM and analyzed parameters are shown in Fig. 1A and 1B. Neither end-diastolic sarcomere length (EDSL), nor the amplitude or velocities of sarcomere shortening-relengthening in PV CM differed from those in LA CM ($p > 0.17$). In both LA and PV CM, absolute and fractional values (FS) of the amplitude of sarcomere shortening positively correlated with the maximum velocity of sarcomere shortening (v_{short}, Fig. 1C).

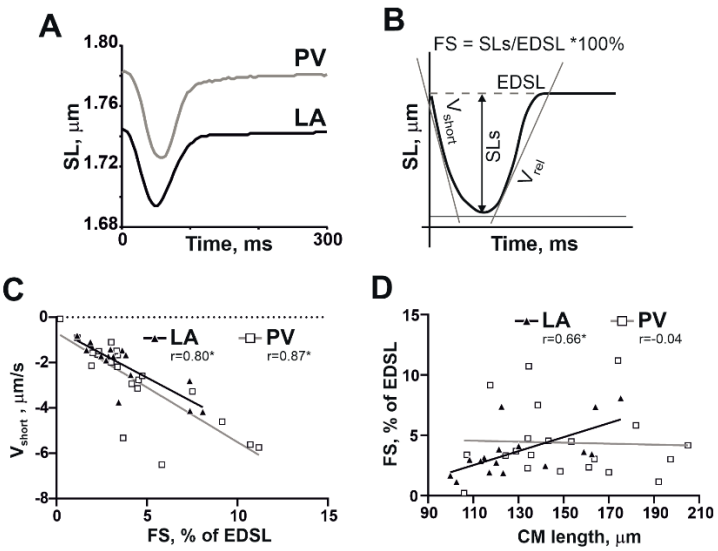


Fig. 1. (A) Representative recordings of sarcomere shortening-relengthening in contracting LA and PV CM. (B) Analyzed parameters derived from the SL change signal: SL – sarcomere length; EDSL – end-diastolic sarcomere length; v_{short} – maximum velocity of sarcomere shortening; FS = sarcomere shortening amplitude/EDSL × 100% – fractional sarcomere shortening amplitude. (C) The correlation of v_{short} with FS. (D) The correlation of FS with CM length

In LA CM, both absolute and fractional sarcomere shortening amplitude increased with an increase in CM length (Fig. 1D). In contrast, in PV CM we did not find any correlation between the characteristics of SL changes and cell morphometry (Fig. 1D).

We then studied the properties of the LA and PV myocardium at the molecular level. We compared the functional properties of myosin from the LA and PV. The sliding velocity of thin filaments at saturating Ca^{2+} concentration over myosin from the PV and LA in the *in vitro* motility assay did not differ ($p > 0.87$) that is the properties of myosin was similar. The cMyBP-C phosphorylation between the PV and LA was not different, while RLC phosphorylation was ~1.7-fold lower in the PV than in the LA ($p = 0.0065$). In the PV, TnT phosphorylation was ~1.4-fold higher ($p = 0.0130$) and TnI phosphorylation ($p = 0.0043$) was ~2.0-fold lower compared to the LA.

It was shown that RLC phosphorylation is involved in the realization of the Frank-Starling and the slow force responses in the atria [4]. Lower levels of RLC phosphorylation in the PV may result in less prominent length-dependent changes in the force of the PV than of the LA. The LA *vs.* PV differences in the levels of TnT and TnI phosphorylation may have complex effects on the contractility of PV and LA CM [5, 6]. A lower phosphorylation level of TnI may lead to the impaired length-dependent activation of myofilaments in PV CM than in LA CM.

Conclusion: The shape of cardiomyocytes is important for contractile function and depends on the strain of the myocardium [7]. In PV CM, there is no correlation between CM morphometric parameters and the amplitude of sarcomere shortening making PV CM more sensitive to volume overload than LA CM. These features, together with the electrophysiological properties of the PV myocardium, might contribute to the occurrence and development of AF.

Funding: The study is supported by the Russian Science Foundation (No. 23-24-00356).

References

1. Roux N., Havet E., Mertl P. The myocardial sleeves of the pulmonary veins: potential implications for atrial fibrillation. *Surg Radiol Anat*. 2004;26:285-289. doi 10.1007/s00276-003-0219-6
2. Lee Y., Cansız B., Kaliske M. Computational modelling of mechano-electric feedback and its arrhythmogenic effects in human ventricular models. *Comput Methods Biomech Biomed Engin*. 2022;25(15):1767-1783. doi 10.1080/10255842.2022.2037573
3. Butova X., Myachina T., Simonova R., Kochurova A., Mukhlynina E., Kopylova G., Shchepkin D., Khokhlova A. The inter-chamber differences in the contractile function between left and right atrial cardiomyocytes in atrial fibrillation in rats. *Front Cardiovasc Med*. 2023;10:1203093. doi 10.3389/fcvm.2023.1203093
4. Kockskämper J., Khafaga M., Grimm M. et al. Angiotensin II and myosin light-chain phosphorylation contribute to the stretch-induced slow force response in human atrial myocardium. *Cardiovasc Res.* 2008;79(4):642-651. doi 10.1093/cvr/cvn126
5. Kumar M., Govindan S., Zhang M. et al. Cardiac Myosin-binding protein C and troponin-I phosphorylation independently modulate myofilament length-dependent activation. *J Biol Chem*. 2015;290(49):29241-29249. doi 10.1074/jbc.M115.686790
6. Warren C.M., Halas M., Goldspink P.H. et al. Truncation of the N-terminus of cardiac troponin I initiates adaptive remodeling of the myocardial proteosome via phosphorylation of mechano-sensitive signaling pathways. *Mol Cell Biochem.* 2022;477(6):1803-1815. doi 10.1007/s11010-022-04414-3
7. Russell B., Curtis M.W., Koshman Y.E., Samarel A.M. Mechanical stress-induced sarcomere assembly for cardiac muscle growth in length and width. *J Mol Cell Cardiol*. 2010;48(5):817-823. doi 10.1016/j.yjmcc.2010.02.016

doi 10.18699/bgrs2024-9.2-28

GlucoGenes®: database of genes and proteins associated with hyperglycemia, hypoglycemia and glucose variability in diabetes

Shishin K.S., Lashin S.A., Klimontov V.V.*

Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

** klimontov@mail.ru*

Key words: diabetes; database; genes; hyperglycemia; hypoglycemia; glucose variability

Motivation and Aim: The search for genes associated with diabetes and glucose fluctuations is important for deciphering the mechanisms of diabetes development and its complications, as well as for identifying promising targets for therapy [1, 2]. The aim of this work was to develop a database for searching and visualizing the information on genes associated with hyperglycemia, hypoglycemia, and both.

Methods and Algorithms: The results of our previous studies analyzing gene networks associated with hyperglycemia, hypoglycemia, and glucose variability (GV) [3–5] were used to create the database. These mentioned studies involved a bioinformatic analysis of the gene networks and text mining of PubMed-indexed publications with the use of the ANDSystem, a bioinformatic tool developed at the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences. The ANDSystem knowledge base contains data about gene regulation, gene associations with diseases, protein-protein interactions, transport pathways, etc., extracted from the Pubmed abstract collection [6, 7]. REST technology was used to control access to the database. The graphical user interface was developed using Vue.js. The database is based on PostgreSQL database management system.

Results: The GlucoGenes® database includes six key tables reflecting genes and corresponding proteins, glucose metabolism disorders (hyperglycemia, hypoglycemia, GV), patterns of associations between genes/proteins and glucose metabolism disorders, references and summary data.

The GlucoGenes® database graphical user interface is available at https://glucogenes.sysbio.ru/genes. It includes the following functional sections:

- Home page: contains general information about GlucoGenes®;
- Catalog of diabetes-associated genes: a gene name, description, and association with glucose disorders;
- Catalog of proteins encoded by the genes contains information about the proteins and their associations with glucose disorders. There are hyperlinks to other catalogs and references.
- Download page.

The search can be performed by gene name, NCBI gene identifier and a kind of glucose disorders.

The database currently contains data on 561 genes associated with glucose disorders and 2115 records describing these relationships.

Conclusion: GlucoGenes® can be accessed at: https://glucogenes.sysbio.ru/genes. It contains information on genes associated with hyperglycemia, hypoglycemia and

increased glucose variability (both hyperglycemia and hypoglycemia). GlucoGenes® may be used in studies of the molecular genetics of diabetes and associated diseases.
Funding: The study was funded by Russian Science Foundation (grant No. 20-15-00057-П).

References

1. Klimontov V.V., Berikov V.B., Saik O.V. Artificial intelligence in diabetology. *Diabetes Mellitus*. 2021;24(2):156-166. doi 10.14341/DM12665
2. Orlov Y.L., Anashkina A.A., Klimontov V.V, Baranova A.V. Medical Genetics, Genomics and Bioinformatics Aid in Understanding Molecular Mechanisms of Human Diseases. *Int J Mol Sci.* 2021;22(18):9962. doi 10.3390/ijms22189962
3. Saik O.V., Klimontov V.V. Bioinformatic Reconstruction and Analysis of Gene Networks Related to Glucose Variability in Diabetes and Its Complications. *Int J Mol Sci.* 2020;21(22):8691. doi 10.3390/ijms21228691
4. Saik O.V., Klimontov V.V. Hypoglycemia, Vascular Disease and Cognitive Dysfunction in Diabetes: Insights from Text Mining-Based Reconstruction and Bioinformatics Analysis of the Gene Networks. *Int J Mol Sci*. 2021;22(22):12419. doi 10.3390/ijms222212419
5. Saik O.V., Klimontov V.V. Gene Networks of Hyperglycemia, Diabetic Complications, and Human Proteins Targeted by SARS-CoV-2: What Is the Molecular Basis for Comorbidity? *Int J Mol Sci*. 2022;23(13):7247. doi 10.3390/ijms23137247
6. Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: An Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst Biol.* 2015;9:S2. doi 10.1186/1752-0509-9-S2-S2
7. Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinform*. 2019;20:34. doi 10.1186/s12859-018-2567-6

9 Симпозиум «Биомедицина, биоинформатика и системная компьютерная биология»

Symposium “Biomedicine, bioinformatics and systems computational biology”

Секция поддержана грантом
Фонда научно-технологического развития
Ханты-Мансийского автономного округа – Югры.
Соглашение № 2023-573-05/2023

doi 10.18699/bgrs2024-9.3-01

Клеточные модели семейной гиперхолестеринемии – новые горизонты борьбы с атеросклерозом

Захарова И.[1]*, Шевченко А.[1], Слепцов А.[3], Назаренко М.[3], Тмоян Н.[2], Зуева А.[1, 4], Арссан А.[1, 4], Елисафенко Е.[1], Шевченко В.[1], Зарубин А.[3], Ежов М.[2], Кухарчук В.[2], Парфёнова Е.[2], Закиян С.[1]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Национальный медицинский исследовательский центр кардиологии имени академика Е.И. Чазова, Москва, Россия*

[3] *Томский национальный исследовательский медицинский центр РАН, Томск, Россия*

[4] *Новосибирский государственный университет, Новосибирск, Россия*

* *zakharova@bionet.nsc.ru*

Ключевые слова: семейная гиперхолестеринемия; атеросклероз; клеточные модели; дифференцировка ИПСК; эндотелиоциты

Мотивация и цель: В настоящее время крайне актуальной задачей является создание клеточных моделей семейной гиперхолестеринемии (СГХС) – наследственного моногенного заболевания, приводящего к атеросклерозу и повышенному риску сердечно-сосудистых патологий. Несмотря на высокую частоту встречаемости (1 на 250 человек для гетерозиготной формы, до 1 на 300 тыс. – 1 млн человек – для гомозиготной) и всемирную озабоченность общественного здравоохранения, эффективность помощи пациентам остается крайне низкой. По данным 2022 года, менее 3 % пациентов в мире, проходящих лечение в связи с СГХС, достигают целевых показателей холестерина липопротеинов низкой плотности. Широкая распространенность, отсутствие общедоступного эффективного лечения и удобных релевантных моделей делают актуальным разработку клеточных моделей СГХС, отвечающих требованиям современной трансляционной медицины. Большинство случаев семейной гиперхолестеринемии вызваны патогенными аллельными вариантами в гене рецептора липопротеинов низкой плотности *LDLR*. Основная роль в патогенезе СГХС принадлежит рецепторам ЛПНП в клетках печени, ответственным за интернализацию холестерина и его последующий метаболизм. Повышенный уровень холестерина ЛПНП в крови также влияет на кровеносные сосуды и сосудистые клетки посредством накопления липидов внутри интимы, вызывая окислительный стресс, рекрутирование воспалительных клеток и локальную продукцию цитокинов. Это приводит к эндотелиальной дисфункции и образованию атером в артериях, корне аорты и клапанах с последующим развитием более тяжелых патологий. Несмотря на то что эндотелий участвует в патогенезе СГХС, мало известно о том, как сосудистые эндотелиальные клетки различаются у людей с нормальными и патогенными аллелями *LDLR*. Целью данной работы является получение изогенной системы линий ИПСК пациентов с СГХС, их релевантных дифференцированных производных и васкуляризированных гепатоорганоидов на их основе.

Методы и алгоритмы: ИПСК получены путем репрограммирования мононуклеарных клеток пациентов с использованием неинтегрирующихся

эписомных векторов. Полученные линии ИПСК пациентов с СГХС охарактеризованы по требованиям Европейского реестра плюрипотентных стволовых клеток человека https://hpscreg.eu/. Плюрипотентность подтверждена с помощью иммунофлуоресцентной окраски антителами к маркерам плюрипотентности, полуколичественной оценки экспрессии генов плюрипотентности методом ПЦР в режиме реального времени, иммунофлуоресцентного анализа способности образовывать производные трех зародышевых листков. Подтверждение наличия аллельных вариантов гена *LDLR* в полученных ИПСК выполнено посредством секвенирования по Сенгеру. Подтверждение генетической идентичности ИПСК и мононуклеарных клеток, из которых они получены, проводилось посредством STR-анализа. Фазирование патологических аллельных вариантов гена *LDLR* выполнено с помощью секвенирования Oxford Nanopore. В ходе работы проведена направленная дифференцировка ИПСК в эндотелиальном, мезенхимальном стромальном и гептоцитарном направлениях, получены выскуляризированные гепатоорганоиды. Характеристика клеток проводилась методами иммунофлуоресцентного анализа и проточной цитометрии. Для характеристики эндотелиоцитов использовался тест на ангиогенез *in vitro*, анализ способности поглощать меченные липопротеины низкой плотности. Секвенирование РНК проводили на платформе Illumina HiSeq 1500. Получение генетически модифицированных изогенных линий ИПСК проводили методом редактирования оснований (base editing) с использованием редакторов цитозиновых оснований BE4 и адениновых оснований ABE7.10.

Результаты: В результате работы получены и охарактеризованы 9 линий ИПСК от пациентов-компаундных гетерозигот по патогенным аллельным вариантам гена *LDLR* с тяжелыми клиническими проявлениями СГХС: c.530C>T/c.1054T>C; c.2141-966_2390-330del/c.1327T>C;
c.1246C>T/c.940+3_940+6del/c.940+3_940+6del. 3 линии зарегистрированы в Европейском реестре плюрипотентных стволовых клеток hPSCreg [1-3]. С помощью нанопорового секвенирования показано, что патогенные аллельные варианты гена *LDLR* в геномах всех пациентов располагаются в транс-положении [4].

В данной работе впервые получены и исследованы эндотелиальные производные ИПСК пациентов с СГХС. Обнаружено, что эндотелиальные производные, полученные из ИПСК, экспрессируют белок LDLR, который в нормальных клетках выявляется преимущественно в зрелой форме [5]. Эндотелиальные клетки, полученные из ИПСК пациентов с СГХС, демонстрируют пониженный уровень зрелого LDLR и сниженную способность к интернализации липопротеинов низкой плотности. Кроме того, эндотелиальные клетки с мутантным *LDLR* обнаруживают специфический профиль транскриптома с пониженной регуляцией генов транспорта монокарбоновых кислот, экзоцитоза и клеточной адгезии, а также с усиленной регуляцией сигнальных путей клеточной секреции и активации лейкоцитов.

ИПСК пациентов с СГХС дифференцированы также в гепато- и мезенхимальные стромальные производные. На их основе с добавлением эндотелиальных производных получены васкуляризированные гепатоорганоиды. В результате секвенирования РНК обнаружено, что в гепатоорганоидах подавлялось большинство генов, связанных с биологическими процессами (регуляция активации клеток/лейкоцитов, позитивная регуляция сигнального пути Wnt,

регуляция роста клеток, клеточный ответ на химический стресс). Среди генов с повышенной экспрессией обнаружены гены, связанные с регуляцией сборки белково-липидного комплекса (GO:0065005, BIN1, SOAT2, MFSD2A).
С использованием CRISPR-Cas9-опосредованого редактирования оснований получены изогенные линии ИПСК пациента-компацндной гетерозиготы c.530C>T/c.1054T>C со скорректированными аллельными вариантами гена *LDLR*. В одной линии произошла коррекция патогенного аллельного варианта c.530C>T до «здорового» c.530C. Таким образом, данная линия после коррекции представляет собой гомозиготу по «здоровой» позиции c.530C и остается гетерозиготой по патогенному аллельному варианту c.1054T>C. В двух других линиях произошла коррекция гетерозиготной позиции c.530C/T до гомозиготы по «больному» c.530C>T варианту. Таким образом, эти две линии представляют собой гомозиготу по «больной» позиции c.530T и остаются гетерозиготами по патогенному аллельному варианту c.1054T>C. Продемонстрировано, что эндотелиальные производные генетически модифицированных ИПСК со скорректированным до «здорового» аллельным вариантом *LDLR* восстанавливают функциональность рецептора и имеют достоверно повышенный уровень интернализации меченной формы липопротеинов низкой плотности.
Выводы: Полученные ИПСК являются первой в России коллекцией ИПСК пациентов с разными формами патологических мутаций гена *LDLR*, представляющей собой доступный практически в неограниченных количествах материал для направленной дифференцировки в релевантные клеточные типы и исследования молекулярно-генетических основ атеросклероза, вызванного СГХС. Полученные результаты указывают на то, что эндотелиальные клетки пациентов с СГХС сами по себе более предрасположены к окислительному стрессу и воспалению, что вместе с повышенным внешним уровнем холестерина может ускорять эндотелиальную дисфункцию, способствуя более быстрому прогрессированию атеросклероза и других сердечно-сосудистых патологий, связанных с СГХС. Генетическая коррекция патогенного аллельного варианта гена *LDLR* восстанавливает функциональность рецептора ЛПНП. Полученные клеточные модели СГХС внесут вклад в исследование заболевания и эффективную разработку таргетных терапевтических препаратов.
Финансирование: Исследование поддержано грантом РНФ (№ 24-15-00346).

Cellular models of familial hypercholesterolaemia: new horizons in the fight against atherosclerosis

Zakharova I.[1]*, Shevchenko A.[1], Sleptcov A.[3], Nazarenko M.[3], Tmoyan N.[2], Zueva A.[1, 4], Arssan A.[1, 4], Elisaphenko E.[1], Shevchenko V.[1], Zarubin A.[3], Ezhov M.[2], Kukharchuk V.[2], Parfyonova Y.[2], Zakian S.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *National Medical Research Centre of Cardiology Named after Academician E.I. Chazov, Ministry of Health of Russian Federation, Moscow, Russia*

[3] *Research Institute of Medical Genetics, Tomsk National Research Medical Centre, RAS, Tomsk, Russia*

[4] *Novosibirsk state university, Novosibirsk, Russia*

* *zakharova@bionet.nsc.ru*

Key words: familial hypercholesterolemia; atherosclerosis; cell models; iPSC differentiation; endothelial cells

Motivation and Aim: Currently, an extremely urgent task is to create cell models of familial hypercholesterolemia (FH), a hereditary monogenic disease that leads to atherosclerosis and an increased risk of cardiovascular pathologies. Despite the high incidence of the disease (1 in 250 people for the heterozygous form, up to 1 in 300 thousand – 1 million people – for the homozygous form) and the worldwide public health concern, the effectiveness of patient care remains extremely low. As of 2022, less than 3 % of patients worldwide being treated for FH achieve LDL cholesterol targets. The pervasive prevalence, lack of publicly available effective treatment, and the absence of convenient relevant models make it imperative to develop cell models of FH that align with the standards of modern translational medicine. The majority of cases of familial hypercholesterolemia are attributable to pathogenic allelic variants in the low-density lipoprotein receptor gene (*LDLR*). The primary role in the pathogenesis of FH is attributed to LDL receptors in liver cells, which are responsible for the internalisation of cholesterol and its subsequent metabolism. Furthermore, elevated levels of LDL cholesterol in the bloodstream also affect blood vessels and vascular cells through the accumulation of intra-intimal lipids, which in turn causes oxidative stress, the recruitment of inflammatory cells, and the local production of cytokines. This results in endothelial dysfunction and the formation of atheromas in the arteries, aortic root and valves, with the subsequent development of more severe pathologies. Although the endothelium has been implicated in the pathogenesis of FH, there is a paucity of knowledge regarding the differences between vascular endothelial cells in individuals with normal and pathogenic *LDLR* alleles. The objective of this study is to establish an isogenic system of induced pluripotent stem cell (iPSC) lines derived from patients with FH, along with their relevant differentiated derivatives and vascularised hepatocyte organoids.

Methods and Algorithms: iPSCs are generated by reprogramming patient-derived mononuclear cells using non-integrating episomal vectors. The resulting iPSC lines from patients with FH were characterised in accordance with the requirements of the European Registry of Human Pluripotent Stem Cells (https://hpscreg.eu/). The pluripotency of the cells was confirmed using immunofluorescent staining with antibodies to pluripotency markers, semi-quantitative assessment of the expression of pluripotency genes using real-time PCR, and immunofluorescent analysis of the ability to form derivatives of the three germ layers. The presence of allelic variants of the *LDLR* gene in the obtained iPSCs was confirmed using Sanger sequencing. The genetic identity of the iPSCs and the mononuclear cells from which they were derived was confirmed using STR analysis. The phasing of pathological allelic variants of the *LDLR* gene was conducted using Oxford Nanopore sequencing. During the course of the study, the directed differentiation of iPSCs was conducted in three distinct directions, namely endothelial, mesenchymal stromal and hepatocyte. The resulting vascularised hepatocyte organoids were obtained. Cell characterisation was conducted via immunofluorescence analysis and flow cytometry. To characterise endothelial cells, an in vitro angiogenesis test was employed, whereby the ability to absorb labelled low-density lipoproteins was analysed. RNA sequencing was conducted on the Illumina HiSeq 1500 platform. Genetically modified isogenic iPSC lines were obtained by base editing using the BE4 cytosine base and ABE7.10 adenine base editors.

Results: The work yielded nine iPSC lines obtained and characterised from patients who were compound heterozygotes for pathogenic allelic variants of the LDLR gene with severe clinical manifestations of FH. These variants included c.530C>T/c.1054T>C; c.2141-966_2390-330del/c.1327T>C; and c.1246C>T/c.940+3_940+6del/c.940+3_940+6del. Three lines are registered in the European Pluripotent Stem Cell Registry (hPSCreg) [1-3]. The use of nanopore sequencing revealed that pathogenic allelic variants of the LDLR gene were located in the trans position in the genomes of all patients [4].

In this study, endothelial derivatives of iPSCs derived from patients with FH were obtained and studied for the first time. It was demonstrated that endothelial derivatives derived from iPSCs express the LDLR protein, which is predominantly expressed in the mature form in normal cells [5]. Endothelial cells derived from iPSCs from patients with FH exhibited reduced levels of mature LDLR and a reduced ability to internalise low-density lipoproteins. Furthermore, endothelial cells with a mutated LDLR exhibit a specific transcriptome profile, characterised by the downregulation of monocarboxylic acid transport, exocytosis, and cell adhesion genes, as well as the upregulation of cellular secretion and leukocyte activation signalling pathways.

Additionally, iPSCs derived from patients with FH can be differentiated into hepatocyte and mesenchymal stromal cell derivatives. Based on these findings, vascularised hepato-organoids were generated by the addition of endothelial derivatives. RNA sequencing revealed that the majority of genes associated with biological processes were suppressed in hepatoorganoids, including those involved in regulating cell/leukocyte activation, positive regulation of the Wnt signalling pathway, regulation of cell growth, and cellular response to chemical stress. Among the genes with increased expression, those associated with the regulation of protein-lipid complex assembly were identified (GO:0065005, BIN1, SOAT2, MFSD2A).

Isogenic iPSC lines from a patient with a heterozygous c.530C>T/c.1054T>C mutation in the LDLR gene were obtained using CRISPR-Cas9-mediated base editing. Allelic variants of the LDLR gene were corrected in these lines. In one line, the pathogenic allelic variant c.530C>T was corrected to the “healthy” c.530C. Consequently, the aforementioned line, following correction, is homozygous for the "healthy" position c.530C and remains heterozygous for the pathogenic allelic variant c.1054T>C. In two additional lines, the heterozygous position c.530C/T was corrected to homozygote for the “sick” c.530C>T variant. Consequently, these two lines are homozygous for the diseased position c.530T and remain heterozygous for the pathogenic allelic variant c.1054T>C. It has been demonstrated that endothelial derivatives of genetically modified iPSCs with the allelic variant of LDLR corrected to a “healthy” one restore the functionality of the receptor and have a significantly increased level of internalisation of the labelled form of low-density lipoprotein.

Conclusion: The obtained iPSCs represent the inaugural collection of iPSCs from patients with various forms of pathological mutations of the LDLR gene in Russia. This material is available in almost unlimited quantities, offering a valuable resource for targeted differentiation into relevant cell types and research into the molecular genetic basis of atherosclerosis caused by FH. The findings indicate that endothelial cells from patients with FH are themselves more susceptible to oxidative stress and inflammation. This, in conjunction with elevated extrinsic cholesterol levels, may accelerate endothelial dysfunction, contributing to a more rapid progression of atherosclerosis and other cardiovascular pathologies associated with FH. The findings indicate that

endothelial cells from patients with FH are themselves more susceptible to oxidative stress and inflammation. This, together with elevated extrinsic cholesterol levels, may accelerate endothelial dysfunction, contributing to more rapid progression of atherosclerosis and other cardiovascular pathologies associated with FH. The correction of a pathogenic allelic variant of the LDLR gene restores the functionality of the LDL receptor. The resulting cell models of FH will contribute to the study of the disease and the development of targeted therapeutic drugs, which will be effective in treating the disease.

Funding: The study is supported by a grant from the Russian Science Foundation (No. 24-15-00346).

Список литературы/References

1. Zakharova I.S. et al. Induced pluripotent stem cell line ICGi037-A, obtained by reprogramming peripheral blood mononuclear cells from a patient with familial hypercholesterolemia due to heterozygous p.Trp443Arg mutations in LDLR. *Stem Cell Res*. 2022;60:102703. doi 10.1016/j.scr.2022.102703
2. Zakharova I.S. et al. Induced pluripotent stem cell line ICGi038-A, obtained by reprogramming peripheral blood mononuclear cells from a patient with familial hypercholesterolemia due to compound heterozygous c.1246C > T/c.940 + 3_940 + 6del mutations in LDLR. *Stem Cell Res.* 2022;60:102702. doi 10.1016/j.scr.2022.102702
3. Zakharova I.S. et al. Induced pluripotent stem cell line ICGi036-A generated by reprogramming peripheral blood mononuclear cells from a patient with familial hypercholesterolemia caused due to compound heterozygous p.Ser177Leu/p.Cys352Arg mutations in LDLR. *Stem Cell Res*. 2022;59:102653. doi 10.1016/j.scr.2022.102653
4. Nazarenko M.S. et al. Calling and phasing of single-nucleotide and structural variants of the ldlr gene using oxford nanopore MinION. *Int J Mol Sci.* 2023;24:4471. doi 10.3390/ijms24054471
5. Zakharova I.S. et al. iPSC-Derived Endothelial Cells Reveal LDLR Dysfunction and dysregulated gene expression profiles in familial hypercholesterolemia. *Int J Mol Sci.* 2024;25:689. doi 10.3390/ijms25020689

doi 10.18699/bgrs2024-9.3-02

Идентификация и установление функциональной значимости однонуклеотидных миссенс-вариантов в генах морфогенеза головного мозга в контексте развития психических и когнитивных расстройств

Карагяур М.Н.[1]*, Примак А.Л.[1], Бозов К.Д.[1], Усатова В.С.[2], Берестовой М.А.[2], Шелег Д.А.[1, 3], Арбатский М.С.[1], Цыганков Б.Д.[1, 3], Ткачук В.А.[1], Нейфельд Е.А.[1, 3]

[1] *Факультет фундаментальной медицины, Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

[2] *Федеральное государственное бюджетное учреждение «Федеральный центр мозга и нейротехнологий» Федерального медико-биологического агентства, Москва, Россия*

[3] *Московский государственный медико-стоматологический университет имени А.И. Евдокимова Министерства здравоохранения Российской Федерации, Москва, Россия*

* *m.karagyaur@mail.ru*

Ключевые слова: морфогенез головного мозга; параноидная шизофрения; эндогенная депрессия; миссенс-мутации; нейральные сфероиды; навигационные молекулы

Мотивация и цель: Нарушение функции и экспрессии генов, вовлеченных в процессы формирования и развития головного мозга, считается одной из возможных причин возникновения психических заболеваний [1–3]. Выявление таких генов и их патологических вариантов, а также установление их функциональной значимости в процессах развития ткани мозга является наиболее актуальной задачей молекулярной психиатрии, поскольку открывает перспективы к диагностике, профилактике, а, возможно, и лечению ряда психических расстройств. Целью данного исследования является идентификация в российской популяции геномных вариантов в генах морфогенеза головного мозга, потенциально ассоциированных с развитием параноидной шизофрении и эндогенной депрессии, с последующим установлением их функциональной значимости на *in vitro* моделях.

Методы и алгоритмы: Для поиска предположительно патологических геномных миссенс-вариантов в генах морфогенеза головного мозга в российской популяции было проведено полноэкзомное секвенирование образцов геномной ДНК пациентов, страдающих параноидной шизофренией (11 человек) и эндогенной депрессией (10 человек). Результаты секвенирования выравнивали на референсный геном человека GRCh37.p13/hg19. Особое внимание уделяли смысловым мутациям в экзонах генов навигационных молекул (ADIPOR1, CDH1-27, CDHR1-4, CD44, DCC, EPHA1-10, EPHB1-6, ERBB2-4, IL6R, IL6ST, ITGA3, ITGAV, ITGB1, NRP1-2, PLAUR, PLXNA1-4, PLXNB1-3 и др.), лигандов (ADIPOQ, CHRD, EFNA1-5, EFNB1-3, IL6, NRG1-4, NTN1-4, PLAU, RELN, SHH, SEMA3A-G и др.), нейротрофических факторов (NGF, BDNF, NTF3, GDNF, VEGFA-D), их рецепторов (NTRK1-3, NGFR, GFRA1, GFRA3, RET), и ассоциированных молекул (PLAT, PLG) – всего 140 генов, которые по данным литературы вовлечены в процессы формирования мозговой ткани.

Распространенность ряда идентифицированных смысловых мутаций (BDNF (rs6265), CDH2 (rs17445840, rs1944294), CDH3 (rs12923655, rs3114409), CDH13 (rs4782724), CDH23 (rs10999947, rs1227051), CDH19/DCHS1 (rs4758443), CDH27/DCHS2 (rs1352714, rs12500437, rs11935573, rs28561984, rs72731014), PLAU (rs2227564) и PLAUR (rs4760)) в популяции пациентов, страдающих шизофренией (102 человека), депрессией (79 человек), а также в популяции здоровых доноров (103 человека) была оценена с помощью модификации аллель-специфичного ПЦР (Amplification-refractory mutation system, ARMS).

Для подтверждения функциональной значимости некоторых из идентифицированных миссенс-мутаций (CDH13, DCHS2, EPHA1, NGF, PLAU и PLXNA3), соответствующие миссенс-мутации будут внесены в индуцированные плюрипотентные клетки здорового донора с помощью технологии редактирования генома CRISPR/Cas9 (редактор оснований). На основе полученных линий иПСК будут собраны нейроглиальные сфероиды.

Результаты: В результате полноэкзомного секвенирования у пациентов, страдающих когнитивными нарушениями, было обнаружено 226 миссенс-мутаций в 79 генах морфогенеза головного мозга (из 140 исследованных). Для ряда мутаций, проанализированных с помощью ARMS, удалось установить достоверно более высокую распространенность в группах пациентов с шизофренией и эндогенной депрессией, чем в группе здоровых доноров. В частности, была установлена ассоциация между частотой встречаемости геномных вариантов rs1944294-T и rs17445840-T (в гене CDH2), rs11935573-G и rs12500437-G/T (в гене DCHS2) и rs1227051-G/A (в гене CDH23), и вероятностью манифестации параноидной шизофрении и эндогенной депрессии [4].

В настоящий момент производится моделирование ряда идентифицированных миссенс-мутаций (CDH13, DCHS2, EPHA1, NGF, PLAU и PLXNA3) в геном индуцированных плюрипотентных клетках здорового донора с помощью модификации технологии редактирования генома CRISPR/Cas9 (редактор оснований). В нейральных сфероидах, сформированных на основе модифицированных клеточных линий, будут выявлены их морфологические, электрофизиологические и транскриптомные особенности.

Выводы: Был выявлен ряд вариантов генов морфогенеза головного мозга, значимо ассоциированных с предрасположенностью к параноидной шизофрении и эндогенной депрессии в российской популяции. Многие из выявленных генетических вариантов в контексте психических расстройств показаны впервые. Функциональная значимость идентифицированных геномных вариантов нуждается в верификации с помощью *in vitro* и *in vivo* моделей. Полученные результаты помогут лучше понять механизмы развития психических нарушений, а в перспективе разработать методы объективной диагностики, профилактики и, возможно, лечения ряда психических заболеваний.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 22-15-00125, https://rscf.ru/project/22-15-00125/.

Identification and establishment of functional significance of single nucleotide missense variants in brain morphogenic genes in the context of psychiatric and cognitive disorders

Karagyaur M.N.[1]*, Primak A.L.[1], Bozov K.D.[1], Usatova V.S.[2], Berestovoy M.A.[2], Sheleg D.A.[1, 3], Arbatsky M.S.[1], Tsygankov B.D.[1, 3], Tkachuk V.A.[1], Neyfeld E.A.[1, 3]

[1] *Faculty of Medicine, Lomonosov Moscow State University, Moscow, Russia*

[2] *Federal Center of Brain Research and Neurotechnologies of the Federal Medical Biological Agency, Moscow, Russia*

[3] *Federal State Budgetary Educational Institution of the Higher Education "A.I. Yevdokimov Moscow State University of Medicine and Dentistry" of the Ministry of Healthcare of the Russian Federation, Moscow, Russia*

** m.karagyaur@mail.ru*

Key words: brain morphogenesis; paranoid schizophrenia; major depressive disorder; missense mutations; neural spheroids; guidance molecules

Motivation and Aim: Aberration in the function and expression of genes involved in the processes of brain morphogenesis is considered to be one of the possible causes of psychiatric diseases [1–3]. The identification of such genes and their pathological variants, as well as the establishment of their functional significance in the processes of brain development, function and dysfunction is one the most urgent task for molecular psychiatry. This knowledge opens prospects for the diagnosis, prevention, and, possibly, treatment of a number of mental disorders. The aim of this study is to identify genomic variants in brain morphogenic genes that are potentially associated with the predisposition to paranoid schizophrenia and major depressive disorder in the Russian population, and establish their functional significance using *in vitro* models of brain tissue development.

Methods and Algorithms: To identify the missense genomic variants in brain morphogenic genes in the Russian population, we performed the whole-exome sequencing of genomic DNA samples of patients suffering from paranoid schizophrenia (11 human) and major depressive disorder (10 human). The sequencing results were aligned to the reference human genome GRCh37.p13/hg19. Special attention was paid to missence-mutations in exons of genes of guidance molecules receptors (ADIPOR1, CDH1-27, CDHR1-4, CD44, DCC, EPHA1-10, EPHB1-6, ERBB2-4, IL6R, IL6ST, ITGA3, ITGAV, ITGB1, NRP1-2, PLAUR, PLXNA1-4, PLXNB1-3, etc.), their ligands(ADIPOQ, CHRD, EFNA1-5, EFNB1-3, IL6, NRG1-4, NTN1-4, PLAU, RELN, SHH, SEMA3A-G, etc.), neurotrophic factors (NGF, BDNF, NTF3, GDNF, VEGFA-D), their receptors (NTRK1-3, NGFR, GFRA1, GFRA3, RET), and associated molecules (PLAT, PLG). In total, 140 genes involved in brain morphogenesis were studied.

The prevalence of several identified missence-mutations (BDNF (rs6265), CDH2 (rs17445840, rs1944294), CDH3 (rs12923655, rs3114409), CDH13 (rs4782724), CDH23 (rs10999947, rs1227051), CDH19/DCHS1 (rs4758443), CDH27/DCHS2 (rs1352714, rs12500437, rs11935573, rs28561984, rs72731014), PLAU (rs2227564) and PLAUR (rs4760)) in the populations of psychiatric patients (schizophrenia – 102 patients, major depressive disorder – 79 patients) and healthy donors (103 human) was

evaluated using amplification-refractory mutation system (ARMS), a modification of allele-specific PCR.

To confirm the functional significance some of the identified missense-mutations (CDH13, DCHS2, EPHA1, NGF, PLAU and PLXNA3) are being modelled in the genome of induced pluripotent stem cells (iPSC) of a healthy donor using CRISPR/Cas9 base editors. The neural spheroids will be obtained from the iPSC lines.

Results: The whole-exome sequencing of DNA samples from the patients who suffered from cognitive disorders revealed 226 missense-mutations in 79 brain morphogenic genes out of 140 genes studied. For a number of mutations analysed using ARMS, we established a significantly higher prevalence in the groups of patients with schizophrenia and major depressive disorder compared to that in the group of healthy donors. In particular, we found that the occurrence of genomic variants rs1944294-T and rs17445840-T (in the CDH2 gene), rs11935573-G and rs12500437-G/T (in the DCHS2 gene) and rs1227051-G/A (in the CDH23 gene) is more frequent in schizophrenia and depression patients [4].

A number of identified missense-mutations (in CDH13, DCHS2, EPHA1, NGF, PLAU and PLXNA3 genes) are being modelled in the genome of iPSCs obtained from a healthy donor using a modification of the CRISPR/Cas9 genome editing system. The morphological and electrophysiological properties, as well as the transcriptome features of the obtained iPSC-derived spheroids will be studied.

Conclusion: We identified a number of brain morphogenic genes variants significantly associated with susceptibility to paranoid schizophrenia and major depressive disorder in the Russian population. Most of the genetic variants in the context of psychiatric disorders were identified for the first time. The functional significance of the detected genomic variants needs to be verified using in vitro and in vivo models. The results obtained will help to reveal the pathogenetic mechanisms of psychiatric disorders, and in the long term to develop methods of objective diagnosis, prevention and, possibly, treatment of a number of psychiatric diseases.

Funding: The study was supported by the Russian Science Foundation grant No. 22-15-00125, https://rscf.ru/project/22-15-00125/.

Список литературы/References

1. Costa M.R. et al. The marginal zone/layer I as a novel niche for neurogenesis and gliogenesis in developing cerebral cortex. *J Neurosci*. 2007;27(42):11376-11388. doi 10.1523/JNEUROSCI.2418-07.2007
2. Meyerink B.L. et al. Ariadne's Thread in the Developing Cerebral Cortex: Mechanisms Enabling the Guiding Role of the Radial Glia Basal Process during Neuron Migration. *Cells*. 2020;10(1):3. doi 10.3390/cells10010003
3. Primak A. et al. Morphogenetic theory of mental and cognitive disorders: the role of neurotrophic and guidance molecules. *Front Mol Neurosci*. 2024;17:1361764. doi 10.3389/fnmol.2024.1361764
4. Karagyaur M. et al. Novel missense variants in brain morphogenic genes associated with depression and schizophrenia. *Front Psychiatry*. 2024;15:1338168. doi 10.3389/fpsyt.2024.1338168

doi 10.18699/bgrs2024-9.3-03

Влияние ингибитора PARP1 на экспрессию генов, ассоциированных с болезнью Хантингтона, в дифференцированных производных ИПСК

Макеева В.С.[1]*, Дырхеева Н.С.[2], Закиян С.М.[1], Малахова А.А.[1]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

* *vladamakkeeva@gmail.com*

Ключевые слова: болезнь Хантингтона; фермент PARP; срединные шипиковые нейроны; ИПСК; ингибиторы

Мотивация и цель: Болезнь Хантингтона (БХ) – аутосомно-доминантное нейродегенеративное заболевание, вызванное экспансией CAG кодонов в первом экзоне гена *HTT* и приводящее к поражению функций и гибели срединных шипиковых нейронов (СШН) полосатого тела мозга пациентов. Для поиска средства терапии широко применяются клеточные модели на основе клеток пациента, благодаря которым имеется возможность разработать пациент-специфичное лекарство, а также изучить особенности влияния генетического фона на патогенез. PARP-ингибиторы, изначально созданные для борьбы с онкозаболеваниями, связанными с мутациями в гене *BRCA*1/2, в последнее время тестируются для терапии нейродегенерации, в частности болезней Альцгеймера, Паркинсона и Хантингтона.

Вследствие развития патогенных процессов в срединных шипиковых нейронах под влиянием мутантной формы хантингтина и хронического нарушения целостности ДНК происходит гиперактивация PARP, детектирующего одноцепочечные разрывы ДНК и привлекающего участников машины репарации путем синтеза меток поли-(АДФ)-рибозы. Повышенная активность приводит к истощению пула НАД+ в клетке, а следовательно, к нарушению энергетического метаболизма клетки и функционирования других НАД-зависимых ферментов. Кроме того, происходит повышение поли-(АДФ)-рибозо-управляемых ферментов, в частности фактора YY1 – который отвечает за экспрессию генов, связанных с развитием стресса ЭПР. Применение ингибиторов должно снизить токсичные для нейронов последствия: так, показано на мышиных моделях болезни Хантингтона R6/2, PARP-ингибиторы способствуют снижению темпов развития симптомов болезни.

Методы и алгоритмы: Получение ИПСК из мононуклеарных клеток периферической крови человека путем репрограммирования с применением временной экспрессии факторов Яманаки. Отбор и характеристика ИПСК с использованием количественной ПЦР для оценки экспрессии факторов плюрипотентности (*OCT4, NANOG, SOX2*), спонтанной дифференцировки и иммунофлуоресцентного анализа на маркеры трех зародышевых листков. Встройка в AAVS1 локус с помощью CRISPR/Cas9 технологии генетических конструкций для экспрессии сенсора стресса ЭПР XBP1-TagRFP. Поиск, на основе анализа литературных данных, генов раннего и позднего начала БХ, изменение

экспрессии которых установлено на клеточных или мышиных моделях БХ, постмортальных образцах пациентов с БХ. Выделение РНК и синтез кДНК. Анализ уровня экспрессии генов-маркеров СШН, стресса ЭПР, болезни Хантингтона.

Результаты: Мы создали клеточную модель на основе ИПСК пациента с 46 CAG повторами в гене HTT путем их дифференцировки в срединные шипиковые нейроны. Многие факты о механизмах болезни установлены с использованием постмортальных образцов мозга пациентов, мышиных моделей и относятся по большей части к постсимптоматической стадии патогенеза. Для изучения ранних стадий болезни клеточные модели являются более подходящим объектом. Однако необходимо учесть, что популяция получаемых нейронов гетерогенна, и включает кроме СШН и другие типы нейронов и глиальных клеток. И потому результаты экспериментов требуется верификация – поэтому проанализировали экспрессию специфичных генов-маркеров СШН (*CTIP2, ARPP21, GABA, DRD1, DRD2*). Кроме того, для контроля мы выбрали гены с дифференциальной экспрессией как на ранних, так и на поздних этапах болезни Хантингтона и проанализировали отклонения их экспрессии на нашей модели.

Для оценки влияния PARP-ингибиторов на развитие патогенеза мы встроили в геном трансгенную конструкцию, экспрессирующую сенсор XBP1-TagRFP, который при развитии стресса ЭПР светит в красном спектре. Кроме того, мы провели анализ изменения экспрессии генов-маркеров стресса ЭПР – *XBP1, ATF4, GRP78, CHOP* – и генов, изменение экспрессии которых показано как на клеточных и мышиных моделях болезни Хантингтона, так и на постмортальных образцах мозга пациентов.

Выводы: В результате мы получили из шести линий ИПСК пациента с БХ культуры нейронов с разной представленностью СШН, которую оценили при помощи количественной ПЦР и иммунофлуоресцентного окрашивания на специфичные маркеры. Оказалось, что не во всех линиях СШН происходит увеличение экспрессии маркеров стресса ЭПР, что также было видно по поведению сенсора XBP1-TagRFP, и также нет полного совпадениях с литературными данными об изменении экспрессии генов, ассоциированных с болезнью Хантингтона.

Тем не менее для линий СШН с повышенным уровнем экспрессии маркеров стресса ЭПР мы смогли показать его снижение после добавления ингибитора PARP Niraparib.

Финансирование: Работа поддержана грантом РНФ № 23-15-00224.

Effect of PARP1 inhibitor on expression of genes associated with Huntington's disease in differentiated iPSC derivatives

Makeeva V. [1]*, Dyrkheeva N. [2], Zakian S. [1], Malakhova A. [1]

[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2]*Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*

**vladamakkeeva@gmail.com*

Key words: Huntington's disease; PARP enzyme; medium spiny neurons; iPSCs; inhibitors

Motivation and objective: Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder caused by the expansion of CAG codons in the first exon of the HTT gene, leading to dysfunction and death of medium spiny neurons (MSN) in the striatum of patients' brains. Cell models based on patient cells are widely used to search for a therapeutic agent, which makes it possible to develop a patient-specific drug and to study the influence of genetic background on pathogenesis.

PARP inhibitors, originally created to combat cancers associated with mutations in the BRCA1/2 gene, have recently been tested for the treatment of neurodegeneration, in particular Alzheimer's, Parkinson's and Huntington's diseases. As a result of the development of pathogenic processes in medium spiny neurons under the influence of a mutant form of huntingtin and chronic disruption of DNA integrity, hyperactivation of PARP occurs, which detects single-stranded DNA breaks and attracts participants of the repair machine by synthesizing poly-(ADP)-ribose labels. Increased activity leads to depletion of the NAD+ pool in the cell, and, consequently, to disruption of the cell's energy metabolism and the functioning of other NAD-dependent enzymes. In addition, there is an increase in poly-(ADP)-ribose-controlled enzymes, in particular, factor YY1, which is responsible for the expression of genes associated with the development of ER stress. The use of inhibitors should reduce the toxic effects on neurons: thus, it has been shown in mouse models of Huntington's disease R6/2, PARP inhibitors help to reduce the rate of development of disease symptoms. Methods and algorithms: Obtaining iPSCs from human peripheral blood mononuclear cells by reprogramming using transient expression of Yamanaka factors. Selection and characterization of iPSCs using quantitative PCR to assess the expression of pluripotency factors (OCT4, NANOG, SOX2), spontaneous differentiation and immunofluorescence analysis for markers of three germ layers. Insertion into the AAVS1 locus using CRISPR/Cas9 technology of genetic constructs for expression of the ER stress sensor XBP1-TagRFP. Search, based on the analysis of literary data, for genes of early and late onset of HD, the change in expression of which was established in cell or mouse models of HD, postmortem samples of patients with HD. Isolation of RNA and synthesis of cDNA. Analysis of the expression level of genes-markers of SCN, ER stress, Huntington's disease.

Results: We created a cell model based on patient iPSCs with 46 CAG repeats in the HTT gene by differentiating them into medium spiny neurons. Many facts about the mechanisms of the disease have been established using postmortem brain samples of patients, mouse models and relate mostly to the post-symptomatic stage of pathogenesis. For studying the early stages of the disease, cellular models are a more suitable object. However, it is necessary to take into account that the population of the obtained neurons is heterogeneous and includes, in addition to SNH, other types of neurons and glial cells. Therefore, the results of the experiments require verification - therefore, we analyzed the expression of specific marker genes of SNH (CTIP2, ARPP21, GABA, DRD1, DRD2). In addition, for control, we selected genes with differential expression both at the early and late stages of Huntington's disease and analyzed deviations in their expression in our model.

To assess the effect of PARP inhibitors on the development of pathogenesis, we integrated into the genome a transgenic construct expressing the XBP1-TagRFP sensor, which shines in the red spectrum during the development of ER stress. In addition, we

analyzed changes in the expression of ER stress marker genes – XBP1, ATF4, GRP78, CHOP – and genes whose expression changes have been shown both in cell and mouse models of Huntington's disease and in postmortem brain samples from patients.
Conclusions: As a result, we obtained neuronal cultures with different levels of ER stress from six HD patient iPSC lines, which were assessed using quantitative PCR and immunofluorescence staining for specific markers. It turned out that not all ER stress lines show an increase in the expression of ER stress markers, which was also evident from the behavior of the XBP1-TagRFP sensor, and there is also no complete agreement with literature data on changes in the expression of genes associated with HD.
However, for ER stress lines with an increased level of expression of ER stress markers, we were able to show a decrease after the addition of the PARP inhibitor Niraparib.
Funding: The work was supported by the Russian Science Foundation grant (No. 23-15-00224).

doi 10.18699/bgrs2024-9.3-04

Получение дермальных фибробластов человека линии НТ 1608 с генетическим нокдауном *YAP1*

Моргун Е.И.[1]*, Черкашина О.Л.[1], Шитова М.С.[1], Мачинская М.А.[1], Воротеляк Е.А.[1, 2]

[1] *Институт биологии развития РАН, Москва, Россия*

[2] *Московский государственный университет имени М.В. Ломоносова, Москва, Россия*

** lady.morgun2016@yandex.ru*

Ключевые слова: YAP; фибробласты; нокдаун

Мотивация и цель: целью нашего исследования было получение дермальных фибробластов человека линии НТ 1608 с генетическим нокдауном *YAP1* для исследования генов-мишеней пути YAP/TAZ.

Методы и алгоритмы: в эксперименте были использованы иммортализованные фибробласты дермы человека линии НТ 1608. Для ингибирования сигнального пути YAP/TAZ разработали плазмиду, экспрессирующую shRNA, нацеленную на мРНК гена *YAP1*. В качестве контроля использовали shRNA Scramble. Далее при помощи клеток HEK293T собирали лентивирусные частицы, содержащие плазмиду, и трансдуцировали целевые клетки. Спустя 3 дня после трансдукции добавили пуромицин для селекции клеток, содержащих плазмиду. После завершения селекции из клеток выделяли РНК, проводили обратную транскрипцию и ПЦР на ген-мишень пути YAP/TAZ – *CCN2*.

Результаты: показано, что экспрессия гена-мишени пути YAP/TAZ – *CCN2* в фибробластах, трансфицированных плазмидой, экспрессирующей анти-*YAP*1 shRNA, была достоверно ниже, чем в контроле ($p \leq 0.05$).

Выводы: в ходе эксперимента были получены дермальные фибробласты человека линии НТ1608 с генетическим нокдауном *YAP1*.

Финансирование: Исследование поддержано РНФ (проект № 21-74-30015).

Obtaining of the HT 1608 human dermal fibroblasts with gene-tic knockdown of *YAP1*

Morgun E.I.[1]*, Cherkashina O.L.[1], Shitova M.S.[1], Machinskaya M.A.[1], Vorotelyak E.A.[1, 2]

[1] *Koltzov Institute of Developmental Biology of Russian Academy of Sciences, Moscow, Russia*

[2] *Lomonosov Moscow State University, Moscow, Russia*

** lady.morgun2016@yandex.ru*

Key words: YAP; fibroblasts; knockdown

Motivation and Aim: the purpose of our study was to obtain human HT 1608 dermal fibroblasts with YAP1 genetic knockdown for the study of target genes of the YAP/TAZ pathway.

Methods and Algorithms: HT 1608 immortalized human dermal fibroblasts were used in the experiment. To inhibit the YAP/TAZ signaling pathway, a plasmid expressing shRNA targeting the mRNA of the *YAP1* gene was developed. shRNA Scramble was used as a control. Next, lentiviral particles containing plasmid were collected using HEK293T cells and the target cells were transduced. 3 days after transduction, puromycin was added to select cells containing plasmid. After the selection was completed, RNA was isolated from the cells, reverse transcription and PCR were performed on the target gene of the YAP/TAZ pathway – *CCN2*.
Results: it was shown that the expression of the target gene of the YAP/TAZ pathway *CCN2* in fibroblasts transfected with a plasmid expressing anti-YAP1 shRNA, was significantly lower than in the control ($p < 0.05$).
Conclusions: during the experiment, human dermal fibroblasts of the HT1608 line with a genetic knockdown of YAP1 were obtained.
Funding: The study is supported by the Russian Science Foundation (grant 21-74-30015).

doi 10.18699/bgrs2024-9.3-05

Трансгенные клеточные линии для изучения взаимодействия тау-белка с микротрубочками и анализа дисфункции митохондрий, вызванной генетическим вариантом c.2013T>G (p.N279K) в гене *MAPT*

Надточий Ю.*, Павлова С., Медведев С.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
** y.nadtochii@g.nsu.ru*

Ключевые слова: ЛВД-17; N279K; тау; MitoTimer

Мотивация и цель: Клеточные модели на основе нейрональных производных индуцированных плюрипотентных стволовых клеток (ИПСК) человека позволяют проводить биохимические, электрофизиологические и омиксные исследования, направленные на раскрытие механизмов патогенеза, реализующихся на клеточном уровне. Лобно-височная деменция (ЛВД) является тяжелым нейродегенеративным заболеванием, одной из причин развития которого являются генетические варианты в гене *MAPT*. Этот ген расположен на 17-й хромосоме и кодирует белок тау, который регулирует сборку и стабилизацию микротрубочек, участвуя в передаче сигнала в ЦНС и аксональном транспорте.
Данная работа посвящена созданию клеточной модели на основе ИПСК, которая позволит изучить механизмы патогенеза при развитии лобно-височной деменции с паркинсонизмом-17, вызванной генетическим вариантом c.2013T>G (p.N279K) в гене *MAPT*, что может ускорить поиск лекарственных средств.
Методы и алгоритмы: Ранее в лаборатории эпигенетики развития ИЦиГ СО РАН были получены линии ИПСК пациента с генетическим вариантом c.2013T>G (rs63750756, p.N279K) в гене *MAPT*. В данные линии ИПСК с помощью CRISPR-Cas9-опосредованной гомологичной рекомбинации в локусе AAVS1 внесли доксициклин-управляемый трансген биосенсора MitoTimer. Затем трансгенные клоны были охарактеризованы для подтверждения статуса плюрипотентности и запущены в направленную дифференцировку в дофаминергические (ДА) нейроны. На терминальной стадии проводилось иммунофлуоресцентное окрашивание на выявление маркёров ДА-нейронов (TH, TUBB3/TUJ1, SOX6, OTX2).
Также для данного исследования проводили трансфекцию клеток линий HEK плазмидными векторами для временной экспрессии TAU-3R и TAU-4R, меченых флуоресцентными белками TagRFP и TagGFP2 для изучения распределения тубулин-ассоциированных белков TAU в цитоплазме клеток.
Результаты: Генетический вариант c.2013T>G (rs63750756, p.N279K) в гене *MAPT* вызывает повышение уровня транскриптов содержащих экзон 10 и, как следствие, повышение уровня экспрессии формы белка тау – TAU-4R. Данная форма белка характеризуется повышенной склонностью к агрегации, что приводит к нарушению аксонального транспорта, митохондриальной дисфункции, что в конечном итоге приводит к гибели клеток. В ходе данного исследования было установлено, что TAU-3R демонстрирует характерный

паттерн распределения в цитоплазме клеток, по-видимому, вызванный ассоциацией с белками цитоскелета. Форма белка TAU-4R распределена более диффузно по цитоплазме клеток, что согласуется с литературными данными.
Были получены линии ИПСК пациента с вариантом p.N279K в гене *MAPT* и здорового донора, несущих доксициклин-управляемый трансген биосенсора MitoTimer, который позволяет изучать биогенез митохондрий в живых клетках в режиме реального времени.
Выводы: Были получены клеточные линии для изучения молекулярно-генетических механизмов ЛВД-17, вызванной мутацией в гене *MAPT*, а также для тестирования потенциальных лекарственных препаратов.
Финансирование: Исследование выполнено при финансовой поддержке Фонда научно-технологического развития Югры в рамках научного проекта № 2023-573-05.

Transgenic cell lines to study the interaction of tau protein with microtubules and to analyze mitochondrial dysfunction caused by the c.2013T>G (p.N279K) genetic variant in the *MAPT* gene

Nadtochy J.*, Pavlova S., Medvedev S.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** y.nadtochii@g.nsu.ru*

Key words: LVD-17; N279K; tau; MitoTimer

Motivation and Aim: Cellular models based on neuronal derivatives of human induced pluripotent stem cells (iPSCs) allow biochemical, electrophysiological, and omics studies aimed at revealing the mechanisms of pathogenesis realized at the cellular level. Frontal temporal dementia (FTD) is a severe neurodegenerative disease, one of the causes of development of which are genetic variants in the MAPT gene. This gene is located on chromosome 17 and encodes a tau protein that regulates microtubule assembly and stabilization, participating in CNS signal transduction and axonal transport.
This work is devoted to the creation of a cell model based on iPSCs, which will allow us to study the mechanisms of pathogenesis in the development of frontal temporal dementia with parkinsonism-17 caused by the genetic variant c.2013T>G (p.N279K) in the MAPT gene, which may accelerate the search for drugs.
Methods and Algorithms: Previously, the laboratory of developmental epigenetics of the Institute of Cytology and Genetics, SB RAS obtained iPSC lines of a patient with the genetic variant c.2013T>G (rs63750756, p.N279K) in the MAPT gene. A doxycycline-guided MitoTimer biosensor transgene was introduced into these hPSC lines by CRISPR-Cas9-mediated homologous recombination at the AAVS1 locus. The transgenic clones were then characterized to confirm pluripotency status and triggered into directed differentiation into dopaminergic (DA) neurons. At the terminal stage, immunofluorescence staining was performed to detect markers of DA neurons (TH, TUBB3/TUJ1, SOX6, OTX2).

For this study, we also transfected HEK cells with plasmid vectors for temporary expression of TAU-3R and TAU-4R labeled with fluorescent proteins TagRFP and TagGFP2 to study the distribution of tubulin-associated TAU proteins in cell cytoplasm.
Results: The genetic variant c.2013T>G (rs63750756, p.N279K) in the MAPT gene causes an increase in the level of transcripts containing exon 10 and, as a consequence, an increase in the level of expression of the TAU-4R form of tau protein. This form of protein is characterized by increased tendency to aggregation, which leads to impaired axonal transport, mitochondrial dysfunction, which ultimately leads to cell death. In this study, we found that TAU-3R exhibits a characteristic distribution pattern in the cytoplasm of cells, apparently caused by association with cytoskeleton proteins. The TAU-4R protein form is distributed more diffusely throughout the cytoplasm of cells, consistent with literature data.
IPSC lines of a patient with the p.N279K variant in the MAPT gene and a healthy donor carrying the doxycycline-guided transgene biosensor MitoTimer, which allows studying mitochondrial biogenesis in living cells in real time, were obtained.
Conclusion: Cell lines were obtained to study the molecular genetic mechanisms of LVD-17 caused by a mutation in the MAPT gene and to test potential drugs.
Funding: The research was carried out with the financial support of the Fund for Scientific and Technological Development of Yugra within the framework of the scientific project No. 2023-573-05.

doi 10.18699/bgrs2024-9.3-06

Исследование роли PDGFRβ в процессе развития коры головного мозга

Целис Суэскун Х.К.[1]*, Хорева Н.С.[1], Тарабыкин В.С.[2]

[1] *Научно-исследовательский институт нейронаук, лаборатория генетики развития мозга, Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия*

[2] *Institute of Cell Biology and Neurobiology, Charité-Universitätsmedizin, Берлин, Германия*

**juancamilo1297@gmail.com*

Ключевые слова: Кора; киназа; нейрогенез; PDGFRβ; in utero электропорация

Мотивация и цель: В процессе развития коры головного мозга, фосфорилирование белков играет значительную роль в сигнальных путях, контролирующих многие клеточные процессы, включая регуляцию клеточного цикла, пролиферацию, дифференцировку, метаболизм и апоптоз. В предыдущих экспериментах *in vitro* с использованием библиотеки ингибиторов киназ, рецептор тромбоцитарного фактора роста бета (PDGFRβ) был идентифицирован как потенциальный регулятор клеточной судьбы развивающихся нейронов [1]. В связи с этим, была проведена серия экспериментов с применением метода in utero электропорации (IUE) для изучения роли PDGFRβ в развитии коры головного мозга и определении клеточной судьбы.

Методы и алгоритмы: Первая серия экспериментов для изучения потенциальной роли PDGFRβ в нейрогенезе была направлена на исследование влияния его повышенной экспрессии (OverExpression, OE) на процессы миграции и определения клеточной судьбы нейронов. С этой целью, нами был создан генетический конструкт pCAG-PDGFRβ-Flag-IRES-eGFP, несущий открытую рамку считывания гена PDGFRβ для экспрессии в нейронах, а также открытую рамку считывания гена eGFP выполняющего функцию репортного сигнала. Валидацию полученного конструкта проводили *in vitro* на первичных смешанных нейрональных культурах после ex utero электропорации (EUE) на стадии e14.5 эмбрионального развития. На 3DIV проводилось иммуноцитохимическое окрашивание прямыми антител antiFlag-Cy3. Наличие в культурах флуорофора Cy3 совместно с eGFP свидетельствовало о накоплении белка PDGFRβ в нейронах.

Дальнейшие эксперименты проводились *in vivo* с применением метода IUE на стадии e14.5 эмбрионального развития до начала спецификации нейрональных клеток верхних слоев с последующим анализом на стадии e18.5 на линии мышей BALB/c. В качестве контроля использовались конструкты с репортными флуоресцентными белками: pCAG-eGFP и pCAG-Venus. После сбора биологических образцов производились фиксация и подготовка тканей с последующим иммуногистохимическим окрашиванием с помощью специфических антител против молекулярных маркеров Satb2 и Ctip2. Анализ послойного распределения eGFP+ клеток, а также оценка изменения доли Satb2+/eGFP+ и Ctip2+/eGFP+ клеток проводили методом конфокальной микроскопии.

Предыдущие эксперименты *in vitro* показывали, что ингибирование PDGFRβ сунитинибом приводит к увеличению доли Satb2+ клеток [1]. Для установления потенциальной роли исследуемой киназы в процессах нейрогенеза, была проведена вторая серия экспериментов по ингибированию (Knockout, KO) PDGFRβ методом РНК интерференции. С этой целью, нами был создан генетический конструкт на основе плазмидного вектора pSuper.neo.retro+GFP, кодирующий shРНК («short-hairpin») специфический к кодирующей последовательности зрелой мРНК исследуемого гена. Валидацию полученного конструкта проводили *in vitro* на культуре клеток Neuro2A. Для этого, трансфицировали клетки одновременно двумя генетическими конструктами – pCAG-PDGFRβ-Flag-IRES-eGFP и pSuper-shRNA-PDGFRb, после чего оценивали накопление Flag-меченного белка PDGFRβ методом Вестерн блотта. Снижение уровня экспрессии PDGFRβ свидетельствовало об эффективной работе созданного генетического конструкта.

Эксперименты *in vivo* с применением метода IUE для KO PDGFRβ с помощью РНК интерференции были проведены аналогично экспериментам по OE. Смесь ДНК для трансфекции нейронов включала в себя pSuper-shRNA-PDGFRb, pCAG-eGFP и pCAG-Venus. В качестве контроля использовался конструкт, кодирующий бессмысленный «scramble» shРНК, а также конструкты с флуоресцентными белками: pCAG-eGFP и pCAG-Venus.

Для исследования влияния KO PDGFRβ на развитие коры головного мозга, была проведена дополнительная серия экспериментов *in vivo* по инактивации гена PDGFRβ с помощью технологии CRISPR-Cas9. С этой целью, были созданы два генетических конструкта, на основе плазмидного вектора pX330-Cas9-hU6, для делеции большой последовательности ДНК гена PDGFRβ. Эффективность работы созданных нами генетических конструктов была проверена с помощью культур клеток HEK293T. Для этого, трансфицировали клетки одновременно двумя генетическими конструктами: pX330-gRNA-PDGFRb несущий последовательность таргетной ДНК против гена исследуемой киназы, а также pEGF-PDGFRb-P несущий кодирующую последовательность флуоресцентного белка eGFP, экспрессия которого нарушена наличием вставки соответствующей кодирующей последовательности гена PDGFRβ. При эффективной работе комплексов Cas9-гРНК, происходит разрезание pEGF-PDGFRb-P с последующим восстановлением экспрессии eGFP. Результаты валидирующих экспериментов оценивали с помощью метода проточной цитометрии.

Эксперименты *in vivo* с применением метода IUE для KO PDGFRβ с помощью CRISPR-Cas9 системы были проведены аналогично экспериментам по OE. Смесь ДНК для трансфекции нейронов включала в себя два pX330-gRNA-PDGFRb, pCAG-eGFP и pCAG-Venus. В качестве контроля использовался конструкт pX330-Cas9-hU6 а также конструкты с флуоресцентными белками: pCAG-eGFP и pCAG-Venus.

Результаты: Результаты экспериментов с повышенной экспрессией PDGFRβ демонстрируют изменения в определении клеточной судьбы электропорированных нейронов. Наблюдалось значительное уменьшение доли Satb2+/eGFP+ нейронов параллельно с увеличением доли Ctip2+/eGFP+ клеток. Анализ послойного распределения электропорированных нейронов показал, что большая доля PDGFRβ OE клеток находилась в нижних слоях коры, что может свидетельствовать о нарушении миграционных механизмов. Данные результаты

подтверждают гипотезу о том, что исследуемая киназа может играть важную роль в определении клеточной судьбы развивающихся нейронов.
Результаты экспериментов IUE по KO PDGFRβ демонстрировали что PDGFRβ участвует в отрицательной регуляции экспрессии Satb2 в развивающихся нейронах. При пониженном уровне экспрессии PDGFRβ наблюдалось значительное увеличение Satb2+/eGFP+ нейронов. Это коррелирует с полученными ранее результатами in vitro, при которых ингибирование PDGFRβ сунитинибом приводило к увеличению количества клеток Satb2+ [1]. При инактивации исследуемой киназы, однако, не были обнаружены значительные изменения в доле Ctip2+/eGFP+ клеток.
Распределение PDGFRβ KO клеток характеризовалось значительным увеличением количества элетропорированных клеток в нижних слоях, а также в некоторых верхних слоях коры головного мозга. Данные результаты демонстрируют, что инактивация PDGFRβ приводит к нарушению цитоархитектуры коры головного мозга, вероятно, посредством нарушения механизмов миграции.
Выводы: Полученные результаты указывают на важную роль киназы PDGFRβ в механизмах определения клеточной судьбы и миграции нейронов во время развития коры головного мозга. Полученные результаты подтверждают роль PDGFRβ в отрицательной регуляции экспрессии Satb2.
Финансирование: Исследование выполнено при финансовой поддержке Министерства науки и высшего образования Российской Федерации (проект № FSWR-2023-0029).

Study of the role of PDGFRβ in the development of the cerebral cortex

Celis Suescun J.C.[1]*, Khoreva N.S.[1], Tarabykin V.S.[2]

[1] *Research Institute of Neuroscience, Laboratory of Genetics of Brain Development, Nizhny Novgorod State University named after. N.I. Lobachevsky, Nizhny Novgorod, Russia*

[2] *Institute of Cell Biology and Neurobiology, Charité-Universitätsmedizin, Berlin, Germany*

** juancamilo1297@gmail.com*

Key words: cerebral cortex; kinase; neurogenesis; PDGFRβ; in utero electroporation

Motivation and Aim: During cortical development, protein phosphorylation plays a significant role in the signaling pathways controlling many cellular processes, including cell cycle regulation, proliferation, differentiation, metabolism, and apoptosis. In previous in vitro experiments using a library of kinase inhibitors, platelet-derived growth factor receptor beta (PDGFRβ) was identified as a potential regulator of cell fate in developing neurons [1]. In this regard, a series of experiments using in utero electroporation (IUE) technique were carried out to study the role of PDGFRβ in cortical development and cell fate determination.
Methods and Algorithms: The first series of experiments to study the potential role of PDGFRβ in neurogenesis was aimed at studying the effect of its increased expression (OverExpression, OE) on the processes of migration and determination of cell fate of neurons. For this purpose, we created the genetic construct pCAG-PDGFRβ-Flag-IRES-eGFP, which carries the open reading frame of the PDGFRβ gene for expression in

neurons, as well as the open reading frame of the eGFP gene, which functions as a reporter signal. Validation of the resulting construct was carried out in vitro on primary mixed neuronal cultures after ex utero electroporation (EUE) at stage E14.5 of embryonic development. At 3DIV, immunocytochemical staining with direct antiFlag-Cy3 antibodies was performed. The presence of the Cy3 fluorophore together with eGFP in the cultures indicated the accumulation of the PDGFRβ protein in neurons.

Further experiments were carried out *in vivo* using the IUE method at stage e14.5 of embryonic development before the start of the specification of neuronal cells of the upper layers, followed by analysis at stage e18.5 in the BALB/c mouse line. Constructs with reporter fluorescent proteins: pCAG-eGFP and pCAG-Venus were used as a control. After collecting biological samples, tissues were fixed and prepared, followed by immunohistochemical staining using specific antibodies against the molecular markers Satb2 and Ctip2. Analysis of the layer-by-layer distribution of eGFP+ cells, as well as assessment of changes in the proportion of Satb2+/eGFP+ and Ctip2+/eGFP+ cells was carried out using confocal microscopy.

Previous in vitro experiments have shown that inhibition of PDGFRβ by sunitinib leads to an increase in the proportion of Satb2+ cells [1]. To establish the potential role of the kinase under study in the processes of neurogenesis, a second series of experiments on inhibition (Knockout) of PDGFRβ by RNA interference was carried out. For this purpose, we created a genetic construct based on the plasmid vector pSuper.neo.retro+GFP, encoding shRNA ("short-hairpin") specific to the coding sequence of the mature mRNA of the gene under study. Validation of the resulting construct was carried out in vitro on Neuro2A cell culture. To do this, cells were simultaneously transfected with two genetic constructs – pCAG-PDGFRβ-Flag-IRES-eGFP and pSuper-shRNA-PDGFRb, after which the accumulation of Flag-tagged PDGFRβ protein was assessed by Western blotting. A decrease in the level of PDGFRβ expression indicated the effective operation of the created genetic construct.

In vivo experiments using the IUE method for KO PDGFRβ using RNA interference were performed similarly to the OE experiments. The DNA mixture for neuronal transfection included pSuper-shRNA-PDGFRb, pCAG-eGFP, and pCAG-Venus. As a control, we used a construct encoding a nonsense "scramble" shRNA, as well as constructs with fluorescent proteins: pCAG-eGFP and pCAG-Venus.

To study the effect of KO PDGFRβ on the development of the cerebral cortex, an additional series of *in vivo* experiments was conducted to inactivate the PDGFRβ gene using CRISPR-Cas9 technology. For this purpose, two genetic constructs were created, based on the plasmid vector pX330-Cas9-hU6, to delete a large DNA sequence of the PDGFRβ gene. The efficiency of the genetic constructs we created was tested using HEK293T cell cultures. To do this, cells were simultaneously transfected with two genetic constructs: pX330-gRNA-PDGFRb, which carries a target DNA sequence against the gene of the kinase under study, and pEGF-PDGFRb-P, which carries the coding sequence of the fluorescent protein eGFP, the expression of which is disrupted by the presence of an insertion of the corresponding coding sequence of the PDGFRβ gene. When Cas9-gRNA complexes operate efficiently, pEGF-PDGFRb-P is cut, followed by restoration of eGFP expression. The results of the validation experiments were assessed using flow cytometry.

In vivo experiments using the IUE method for KO PDGFRβ using the CRISPR-Cas9 system were performed similarly to the OE experiments. The DNA mixture for neuronal transfection included two pX330-gRNA-PDGFRb, pCAG-eGFP and pCAG-Venus. The

pX330-Cas9-hU6 construct as well as constructs with fluorescent proteins: pCAG-eGFP and pCAG-Venus were used as a control.

Results: Results from PDGFRβ overexpression experiments demonstrate changes in cell fate determination of electroporated neurons. There was a significant decrease in the proportion of Satb2+/eGFP+ neurons in parallel with an increase in the proportion of Ctip2+/eGFP+ cells. Analysis of the layer-by-layer distribution of electroporated neurons showed that a large proportion of PDGFRβ OE cells were located in the lower layers of the cortex, which may indicate a disruption of migratory mechanisms. These results support the hypothesis that the kinase under study may play an important role in determining the cell fate of developing neurons.

Results from IUE PDGFRβ KO experiments demonstrated that PDGFRβ is involved in the negative regulation of Satb2 expression in developing neurons. When PDGFRβ expression was reduced, a significant increase in Satb2+/eGFP+ neurons was observed. This correlates with previous in vitro results in which inhibition of PDGFRβ sunitinib led to an increase in the number of Satb2+ cells [1]. Upon inactivation of the kinase under study, however, no significant changes were detected in the proportion of Ctip2+/eGFP+ cells.

The distribution of PDGFRβ KO cells was characterized by a significant increase in the number of eletroporated cells in the lower layers, as well as in some upper layers of the cerebral cortex. These results demonstrate that inactivation of PDGFRβ leads to disruption of the cytoarchitecture of the cerebral cortex, likely through disruption of migration mechanisms.

Conclusion: Our results indicate an important role for PDGFRβ kinase in cell fate determination and neuronal migration during cortical development. Our results support the role of PDGFRβ in the negative regulation of Satb2 expression.

Funding: The study was carried out with financial support from the Ministry of Science and Higher Education of the Russian Federation (project No. FSWR-2023-0029).

Список литературы/References

1 Ambrozkiewicz M.C., Bessa P., Salazar-Lázaro A., Salina V., Tarabykin V. Satb2$^{Cre/+}$ mouse as a tool to investigate cell fate determination in the developing neocortex. *J Neurosci Methods*. 2017;291:113-121

doi 10.18699/bgrs2024-9.3-07

Изучение влияния делеции HAR, расположенных в гене *CNTN6*, на ранние этапы нейрогенеза человека с помощью церебральных органоидов

Чвилёва А.[1]*, Юнусова А.[2], Пристяжнюк И.[2], Рыжкова А.[2], Смирнов А.[2], Белокопытова П.[1, 2], Шнайдер Т.[2]

[1] *Новосибирский государственный университет, Новосибирск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

* *a.chvileva@g.nsu.ru*

Ключевые слова: эмбриональное развитие; нейрогенез; церебральные органоиды; регуляторные последовательности; умственная отсталость; HAR; ИПСК; *CNTN6*

Мотивация и цель: Нарушение умственного развития считается одним из распространенных расстройств психики. Умственная отсталость встречается у 3 % родившихся детей и имеет высокий уровень сопутствующих нейропсихиатрических патологий. Причины подобных нарушений часто скрываются в наследственных факторах. Показано, что до 25 % этих проблем связаны с copy number variations (CNVs) различных генов. Однако на данный момент нет четкого описания генетической этиологии нарушения когнитивных функций.

Объектом нашего исследования является ген-кандидат для нарушений умственного развития – *CNTN6*. В настоящее время описано несколько десятков пациентов с умственной отсталостью, у которых были обнаружены CNVs данного гена [1–3]. Интересно, что большинство обнаруженных мутаций представлены делециями и дупликациями крупных размеров, однако точечная замена была обнаружена лишь в одном клиническом случае [4]. В связи с этим нами было выдвинуто предположение о наличии функциональных участков в локусе *CNTN6*, изменение в работе которых может привести к нарушениям ранних этапов развития мозга человека.

Аннотированных регуляторных элементов в данном локусе не было обнаружено, однако были выявлены два региона, быстро эволюционировавших у человека – human accelerated region (HAR) [5]. Эти участки ДНК часто являются энхансерами генов, вовлеченных в развитие мозга человека. Также на данный момент известны мутации в HAR, изменяющие экспрессию генов-кандидатов для различных психических расстройств [6]. Цель данной работы – исследование эффекта делеции HAR, расположенных в локусе *CNTN6*, на ранние этапы нейрогенеза человека с помощью церебральных органоидов (ЦО) [7].

Результаты: Для выполнения данной цели с помощью системы CRISPR/Cas9 были получены линии ИПСК с гомозиготными делециями районов, содержащих HAR. Для полученных линий была подтверждена их способность поддерживать плюрипотентное состояние *in vitro* по экспрессии факторов плюрипотентности и наличию потенциала к дифференцировке в три зародышевых листка. Также линии ИПСК были проверены на отсутствие хромосомных аберраций.

Полученные линии были дифференцированы в ЦО. При дифференцировке была выявлена тенденция к уменьшению размеров органоидов с делецией HAR по сравнению с контролем. Исследование внутренней организации ЦО позволило выявить, что на 20-й день дифференцировки в органоидах с делецией HAR достоверно уменьшается толщина и площадь желудочкоподобных структур. Иммуногистохимический анализ на маркер плотных межклеточных контактов ZO1 показал значительные уменьшения апикальной мембраны желудочкоподобных структур органоидов с делецией, что может говорить о нарушении процесса самоорганизации клеток радиальной глии. Этот тип клеток играет важнейшую роль в процессе развития данных структур. Одним из основных маркеров для клеток радиальной глии считается белок PAX6. Ранее в нашей лаборатории было установлено, что наличие делеции гена CNTN6 приводит нарушению клеточной локализации белка и уменьшению его количества в ЦО. Однако в органоидах с делецией HAR подобный фенотип не наблюдался (рис. 1). Иммуногистохимический анализ показал нарушение распределения PAX6 сигнала, однако достоверных изменений доли площади PAX6-позитивного сигнала на срезе ЦО не было обнаружено.

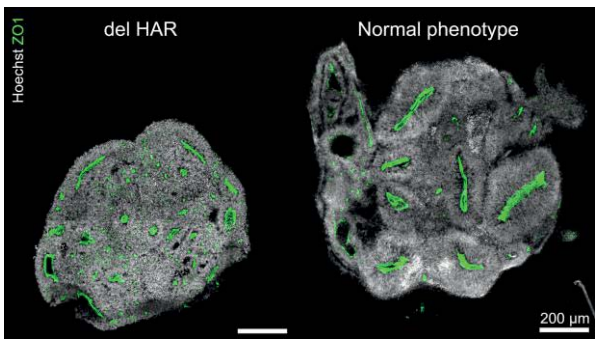


Рис. 1. Иммуноокрашивание контрольных церебральных органоидов (Normal phenotype) и органоидов с делецией (del HAR) с использованием антител против ZO1. Масштабная линейка – 200 мкм

Выводы: Таким образом, делеция HAR в гене *CNTN6* приводит к уменьшению размеров органоидов и нарушениям формирования желудочкоподобных структур, что указывает на участие данной последовательности в регуляции раннего этапа нейрогенеза человека *in vitro*. Однако данная регуляция никак не связана с PAX6-опосредованным контролем дифференцировки клеток радиальной глии.

Финансирование: Работа выполнена при поддержке РНФ (проект № 24-24-00447).
Благодарности: Культивирование линий ИПСК проводили на базе ЦКП «Коллекция плюрипотентных культур клеток человека и млекопитающих общебиологического и биомедицинского направления» ИЦиГ СО РАН. Микроскопический анализ проведен на базе ЦКП микроскопического анализа биологических объектов ИЦиГ СО РАН.

Investigating the effect of HAR deletion located in the *CNTN6* gene on early stages of human neurogenesis on the cerebral organoids model

Chvileva A.[1]*, Yunusova A.[2], Pristyazhnyuk I.[2], Ryzhkova A.[2], Smirnov A.[2], Belokopytova P.[1, 2], Shnaider T.[2]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *a.chvileva@g.nsu.ru*

Key words: embryonic development; neurogenesis; cerebral organoids; regulatory sequences; mental retardation; HAR; iPSCs; *CNTN6*

Motivation and Aim: Mental retardation is one of the most common mental disorders. Intellectual disability occurs in 3 % of born children and has a high rate of associated neuropsychiatric pathologies. The causes of such disorders often lie in hereditary factors. It has been shown that up to 25 % of these problems are associated with copy number variations (CNVs) of various genes. However, there is currently no clear description of the genetic etiology of cognitive impairment.

The object of our study is a candidate gene for mental retardation disorders – *CNTN6*. Currently, several dozens of patients with mental retardation in which CNVs of this gene have been detected have been described [1–3]. Interestingly, large deletions and duplications represent most of the associated mutations, but a point substitution was found in only one clinical case [4]. Therefore, we hypothesized the presence of functional regions in the *CNTN6* locus, the alteration of which may lead to abnormalities in the early stages of human brain development.

Annotated regulatory elements were not found at this locus, but two regions that evolved rapidly in humans – human accelerated region (HAR) – were identified [5]. These DNA regions may function as enhancers of genes involved in human brain development. Mutations in HAR are also known to alter the expression of candidate genes for various psychiatric disorders [6]. This work aims to study the effect of HAR deletion located in the *CNTN6* locus on the early stages of human neurogenesis on the cerebral organoid model (CO) [7].

Results: To realize this goal, hiPSC lines with homozygous deletions of HAR were obtained using the CRISPR/Cas9 system. Their ability to maintain the pluripotent state *in vitro* was confirmed by the expression of pluripotency factors and the potential for differentiation into three germ sheets. iPSC lines were also tested for the absence of chromosomal aberrations.

The obtained lines were differentiated into COs. The differentiation revealed a tendency to decrease the size of organoids with HAR deletion compared to the control. Analysis of the internal organization of the COs revealed that on day 20 of differentiation, the thickness and area of ventricle-like structures were significantly reduced in mutant organoids. Immunohistochemical analysis for the tight junction-associated proteins showed significant reductions in the apical membrane of ventricle-like structures of organoids with HAR deletion (Fig. 1), which may indicate impaired self-organization of radial glial cells.

This cell type plays a crucial role in the process of these structures' development. One of the key markers for radial glial cells is considered to be PAX6 protein. In our previous study it was found that the presence of CNTN6 gene deletion leads to the disruption of cellular localization of the protein and a decrease in its quantity in the COs. However, no such phenotype was observed in organoids with HAR deletion. Immunohistochemical analysis showed the distribution disorder of PAX6 signal; though, no significant changes in the area fraction of PAX6-positive signal were detected on the slice of COs.

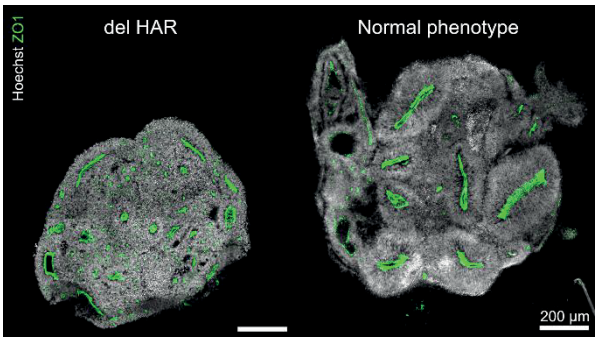


Fig. 1. Immunostaining of control cerebral organoids (Normal phenotype) and organoids with deletion (del HAR) using antibodies against ZO1. The scale bar – 200 µm

Conclusion: Thus, HAR deletion in the *CNTN6* gene leads to decreased organoid size and impaired ventricle-like structure formation, indicating that this region is involved in the regulation of the early stage of human neurogenesis *in vitro*. However, this regulation is unrelated to PAX6-mediated control of radial glial cell differentiation.

Funding: This study was supported by the Russian Science Foundation (project No. 24-24-00447).

Acknowledgement: iPSCs derivation and differentiation was performed at the Collective Center of ICG SB RAS ‘‘Collection of Pluripotent Human and Mammalian Cell Cultures for Biological and Biomedical Research’’. Microscopy assays were performed at the Center of Collective Use for Microscopic Analysis of Biological Objects (ICG SB RAS, Novosibirsk, Russia).

Список литературы/References

1. Kashevarova A.A. et al. Single gene microdeletions and microduplication of 3p26.3 in three unrelated families: *CNTN6* as a new candidate gene for intellectual disability. *Mol Cytogenet*. 2014;7:97
2. Hu J. et al. *CNTN6* copy number variations in 14 patients: a possible candidate gene for neurodevelopmental and neuropsychiatric disorders. *J Neurodev Disord.* 2015;7:26
3. Tassano E. et al. Clinical and molecular characterization of two patients with CNTN6 copy number variations. *Cytogenet Genome Res.* 2018;156(3):144-149
4. García-Ortiz J.E. et al. Case Report: Whole exome sequencing unveils an inherited truncating variant in CNTN6 (p. Ser189Ter) in a mexican child with autism spectrum disorder. *J Autism Dev Disord*. 2020;50:2247-2251
5. Pollard K.S. et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*. 2006;443(7108):167-172
6. Doan R.N. et al. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell*. 2016;167(2):341-354
7. Lancaster M.A. et al. Cerebral organoids model human brain development and microcephaly. *Nature*. 2013;501(7467):373-379

doi 10.18699/bgrs2024-9.3-08

Влияние вариантов c.1977G>A в гене MYH7 и c.1543_1545delAAC в гене MYBPC3 на развитие гипертрофической кардиомиопатии

Шульгина А.*, Павлова С., Проняева К., Закиян С., Дементьева Е.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
** ange.shulgina@yandex.ru*

Ключевые слова**:** гипертрофическая кардиомиопатия; генетический вариант с неясным клиническим значением; CRISPR/Cas9; индуцированные плюрипотентные стволовые клетки; кардиомиоцит

Мотивация и цель: Гипертрофическая кардиомиопатия (ГКМП) – одна из самых распространенных сердечно-сосудистых патологий. Заболевание характеризуется нарушением структуры саркомеров кардиомиоцитов и увеличением их размера, гипертрофией стенок левого желудочка, прогрессирующей сердечной недостаточностью, высоким риском аритмий и внезапной сердечной смерти. В 60% случаев заболевание является наследственным. На сегодняшний день известно более 1400 генетических вариантов, связанных с ГКМП, в генах, которые по большей части кодируют саркомерные белки [1]. Из этих генетических вариантов примерно 80% приходится на такие гены, как *MYH7* и *MYBPC3* [2]. Однако не для всех описанных генетических вариантов подтверждена способность влиять на развитие патологии. Ранее нами были выявлены два пациента, страдающих ГКМП, с вариантами с неясным клиническим значением c.1977G>A и c.1543_1545delAAC в генах *MYH7* и *MYBPC3* соответственно [3].
Целью данной работы является исследование патогенетического вклада вариантов c.1977G>A и c.1543_1545delAAC в генах *MYH7* и *MYBPC3* в развитие ГКМП.
Методы и алгоритмы: Для выяснения клинического значения вариантов c.1977G>A и c.1543_1545delAAC в генах *MYH7* и *MYBPC3* данные мутации были внесены в линию индуцированных плюрипотентных стволовых клеток (ИПСК) здорового донора K7-4Lf [4], а также исправлены в ранее полученных пациент-специфичных линиях ИПСК [5-6]. Редактирование генов *MYH7* и *MYBPC3* в линиях ИПСК осуществляли с помощью системы CRISPR/Cas9 и одноцепочечных донорных олигонуклеотидов, содержащих нуклеотидные последовательности 18-го экзона гена *MYH7* или 17-го экзона гена *MYBPC3*. Донорные олигонуклеотиды для внесения мутаций содержали замену c.1977G>A или делецию c.1543_1545delAAC, донорные олигонуклеотиды для исправления мутаций соответствовали нуклеотидной последовательности здорового донора. Система CRISPR/Cas9 использовалась в виде рибонуклеопротеиновых комплексов, sgRNA:Cas9_NLS, которые вместе с донорными олигонуклеотидами доставлялись в клетки с помощью электропорации. Полученные после электропорации ИПСК субклонировали. Анализ событий в целевых районах в сублонах ИПСК проводили секвенированием по Сэнгеру. Линии ИПСК с внесенными и исправленными мутациями c.1977G>A и c.1543_1545delAAC в генах *MYH7* и *MYBPC3* стандартными методами были проверены на экспрессию маркеров

плюрипотентного состояния, способность дифференцироваться в производные трех зародышевых листков и сохранение нормального кариотипа. Наличие нецелевой активности системы CRISPR/Cas9 в полученных линиях ИПСК проверялось секвенированием по Сэнгеру пяти наиболее вероятных предсказанных нецелевых сайтов системы CRISPR/Cas9. Направленная дифференцировка линий ИПСК с внесенными мутациями c.1977G>A и c.1543_1545delAAC в гены *MYH7* и *MYBPC3*, пациент-специфичных линий ИПСК и линий ИПСК здоровых доноров в кардиомиоциты осуществлялась по протоколу основанному на модуляции сигнального пути Wnt [7]. Размер полученных кардиомиоцитов оценивался с помощью иммунофлуоресцентного окрашивания и последующего подсчета их площадей в программе ImageJ. Для кардиомиоцитов с внесенной мутацией c.1977G>A в ген *MYH7* также проведено исследование митохондриального дыхания с использованием технологии Seahorse (Agilent).

Результаты: В результате редактирования генов *MYH7* и *MYBPC3* линии ИПСК здорового донора и пациент-специфичных линий ИПСК с помощью системы CRISPR/Cas9 были получены одна линия ИПСК, гетерозиготная по внесенной мутации c.1977G>A в ген *MYH7*, три линии ИПСК, гомозиготные по внесенной мутации c.1543_1545delAAC в ген *MYBPC3*, две линии ИПСК с исправленной мутацией c.1977G>A в гене *MYH7* и три линии ИПСК с исправленной мутацией c.1543_1545delAAC в гене *MYBPC3*. Все полученные линии ИПСК сохраняли свои плюрипотентные свойства: экспрессию маркеров плюрипотентного состояния и способность дифференцироваться в производные трех зародышевых листков – и нормальный кариотип. В полученных линиях ИПСК также не было обнаружено нецелевой активности системы CRISPR/Cas9. Было показано, что площади кардиомиоцитов, полученных в результате направленной дифференцировки линий ИПСК с внесенными вариантами c.1977G>A и c.1543_1545delAAC в гены *MYH7* и *MYBPC3* и пациент-специфичных линий ИПСК, достоверно увеличены по сравнению с площадями кардиомиоцитов, полученных при дифференцировке линий ИПСК здоровых доноров. Также базовый уровень потребления кислорода в кардиомиоцитах с внесенным генетическим вариантом c.1977G>A в ген *MYH7* был значительно ниже относительно кардиомиоцитов, полученных из линии ИПСК здорового донора, использованной для внесения мутации.

Выводы: С использованием системы CRISPR/Cas9 была создана панель изогенных линий ИПСК для изучения патогенетического вклада вариантов c.1977G>A и c.1543_1545delAAC в генах *MYH7* и *MYBPC3* в развитие ГКМП. В кардиомиоцитах, полученных из линий ИПСК с внесенными мутациями c.1977G>A и c.1543_1545delAAC в гены *MYH7* и *MYBPC3*, воспроизводится такой важный признак ГКМП, как увеличенный размер. Кроме того, в кардиомиоцитах с внесенной мутацией c.1977G>A в ген *MYH7* обнаружено снижение митохондриального дыхания. Данные результаты свидетельствуют о патогенности вариантов c.1977G>A и c.1543_1545delAAC в генах *MYH7* и *MYBPC3*.

Финансирование: Исследование поддержано РНФ (проект № 22-15-00271).

The effect of variants c.1977G>A in *MYH7* and c.1543_1545delAAC in *MYBPC3* on the development of hypertrophic cardiomyopathy

Shulgina A.*, Pavlova S., Pronyaeva K., Zakian S., Dementyeva E.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** ange.shulgina@yandex.ru*

Key words: hypertrophic cardiomyopathy; genetic variant of unknown significance; CRISPR/Cas9; induced pluripotent stem cells; cardiomyocyte

Motivation and Aim: Hypertrophic cardiomyopathy (HCM) is one of the most common cardiovascular pathologies. The disease is characterized by disruption of sarcomere structure of cardiomyocytes and their enlarged size, hypertrophy of left ventricle walls, progressive heart failure, a high risk of arrhythmia and sudden cardiac death. In 60% of cases, the disease is inherited. To date more than 1400 HCM-associated genetic variants in genes encoding predominately sarcomeric proteins have been described [1]. Among the variants, about 80% were found in *MYH7* and *MYBPC3* [2]. However, capacity to have impact on pathology development is confirmed not for all the genetic variant. We previously identified two HCM patients carrying variants of unknown significance – c.1977G>A and c.1543_1545delAAC in *MYH7* and *MYBPC3*, respectively [3].
The aim of this work is studying pathogenetic contribution of c.1977G>A and c.1543_1545delAAC variants in *MYH7* and *MYBPC3* to HCM development.
Methods and Algorithms: To find out clinical significance of c.1977G>A and c.1543_1545delAAC variants in *MYH7* and *MYBPC3*, the mutations were introduced in induced pluripotent stem cells (iPSCs) of healthy donor, the K7-4Lf line [4] and were corrected in patient-specific iPSC lines generated previously [5-6]. *MYH7* and *MYBPC3* editing in the iPSC lines was performed with CRISPR/Cas9 and single-stranded donor oligonucleotides contained nucleotide sequences of *MYH7* exon 18 or *MYBPC3* exon 17. Donor oligonucleotides for introducing mutations comprised c.1977G>A substitution or c.1543_1545delAAC deletion, donor oligonucleotides for mutation correction corresponded to nucleotide sequences of healthy donor. CRISPR/Cas9 was used in the form of ribonucleoprotein complexes, sgRNA:Cas9_NLS, that were delivered to cells together with the donor oligonucleotides by electroporation. The iPSCs obtained after electroporation were subcloned. Analysis of events in the target regions in the iPSC subclones was carried out using Sanger sequencing. iPSC lines with introduced and corrected c.1977G>A and c.1543_1545delAAC mutations in *MYH7* and *MYBPC3* were checked for expression of pluripotency state markers, capacity to be differentiated into derivatives of three germ layers, and retaining normal karyotype by a standard set of methods. Off-target CRISPR/Cas9 activity in the iPSC lines was verified by Sanger sequencing of top-5 predicted CRISPR/Cas9 off-target sites. Directed differentiation of the iPSC lines with introduced c.1977G>A and c.1543_1545delAAC mutations in *MYH7* и *MYBPC3*, the patient-specific iPSC lines and iPSC lines of healthy donors into cardiomyocytes was performed according to the protocol based on the Wnt signaling pathway modulation [7]. The size of the cardiomyocytes was evaluated using immunofluorescence staining followed by cardiomyocyte area calculation in ImageJ.

For cardiomyocytes with introduced c.1977G>A mutation in *MYH7*, mitochondrial respiration was also examined using the Seahorse (Agilent) technology.

Results: As a result of *MYH7* and *MYBPC3* editing of the iPSC lines of healthy donor and patient-specific iPSC lines using CRISPR/Cas9, one iPSC line heterozygous at introduced c.1977G>A mutation in *MYH7*, three iPSC lines homozygous at introduced c.1543_1545delAAC mutation in *MYBPC3*, two iPSC lines with corrected c.1977G>A mutation in *MYH7*, and three iPSC lines with corrected c.1543_1545delAAC mutation in *MYBPC3* were generated. All the iPSC lines retained their pluripotent properties – expression of the pluripotency state markers and capacity to be differentiated into derivatives of three germ layers – and normal karyotype. Off-target CRISPR/Cas9 activity was not observed in the iPSC lines. It was shown that areas of cardiomyocytes derived from the iPSC lines with introduced c.1977G>A and c.1543_1545delAAC variants in *MYH7* and *MYBPC3* and patient-specific iPSC lines were significantly increased in comparison with those derived from the iPSC lines of healthy donors. Basic oxygen consumption rate in cardiomyocytes with introduced c.1977G>A variant in *MYH7* was decreased compared to cardiomyocytes derived from the iPSC line of healthy donor used for introducing the mutation.

Conclusion: A panel of isogenic iPSC lines for studying pathogenetic contribution of c.1977G>A and c.1543_1545delAAC variants in *MYH7* and *MYBPC3* to HCM development has been generated using CRISPR/Cas9. An important HCM feature, enlarged cardiomyocyte size, was reproduced in cardiomyocytes derived from the iPSC lines with introduced c.1977G>A and c.1543_1545delAAC mutations in *MYH7* and *MYBPC3*. A decrease in mitochondrial respiration was also revealed in cardiomyocytes with introduced c.1977G>A mutation in *MYH7*. The results testify in favor of pathogenicity of c.1977G>A and c.1543_1545delAAC variants in *MYH7* and *MYBPC3*.

Funding: The study was supported by Russian Science Foundation (project No. 22-15-00271).

Список литературы/References

1. Vatutin N.T. et al. Hypertrophic cardiomyopathy: genetic alterations, pathogenesis and pathophysiology. *Russian Journal of Cardiology*. 2014;5(109):35-42
2. Akhtar M., Elliott P. The genetics of hypertrophic cardiomyopathy. *Global Cardiology Science and Practice*. 2018;71(3):132-138
3. Dementyeva E.V. et al. Genetic analysis of patients with hypertrophic cardiomyopathy. *Genes and Cells*. 2020;15(3):68-73
4. Malakhova A.A. et al. Generation of induced pluripotent stem cell lines ICGi021-A and ICGi022-A from peripheral blood mononuclear cells of two healthy individuals from Siberian population. *Stem Cell Res*. 2020;48:101952
5. Dementyeva E.V. et al. Generation of two clonal iPSC lines, ICGi019-A and ICGi019-B, by reprogramming peripheral blood mononuclear cells of a patient suffering from hypertrophic cardiomyopathy and carrying a heterozygous p.M659I mutation in *MYH7*. *Stem Cell Res*. 2020;46:101840
6. Dementyeva E.V. et al. Generation of an induced pluripotent stem cell line, ICGi029-A, by reprogramming peripheral blood mononuclear cell of a patient suffering from hypertrophic cardiomyopathy and carrying a heterozygous p.N515del mutation in *MYBPC3*. *Stem Cell Res*. 2021;53: 102344
7. Burridge P.W. et al. Chemically defined generation of human cardiomyocytes. *Nat. Methods*. 2014;11:855-860

doi 10.18699/bgrs2024-9.3-09

The approaches for CRISPR/Cas9 regulation on the guide RNA level

Novopashina D.S.[1, 2]*, Sakovina L.V.[1, 2], Gorlenko E.S.[1, 2], Dolzhikova O.A.[1, 2], Meschaninova M.I.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*

[2] *Novosiborsk State University, Novosibirsk, Russia*

* *danov@niboch.nsc.ru*

Key words: CRISPR/Cas9 system, guide RNA, regulation, photosensitive linkers, aptamers

Motivation and Aim: The increase of effectivity and accuracy of gene editing system is the actual task of modern molecular biology and gene engineering. Design of regulated CRISPR/Cas systems and development of approaches for controlling their activity by physico-chemical stimuli are especially interesting in the context of gene editing. Among such approaches, activity regulation on the level of guide RNAs by introducing chemical modifications, changing the sequences or adding the blocking oligonucleotide are seem to be very promising. The aim of this research is design and testing of CRISPR/Cas9 system that can be regulated by irradiation and small molecules.

Methods and Algorithms: For the synthesis of guide RNA, we used solid phase automatic phosphoramidate method and corresponding modified or non-nucleotide synthons. Two variants of regulation were tested in model system: irradiation by the light with definite wavelength and allosteric regulation by theophylline.

Results: Linear and cyclic photocleavable guide RNA and blocking oligonucleotide were synthetized. Linear crRNAs and sgRNAs contained the photocleavable linkers which permit to destroy guide RNA in concrete moment, thus inhibiting DNA cleavage by Cas9 nuclease. Cyclic crRNAs do not fit well for CRISPR/Cas9 system, therefore before irradiation the DNA cleavage is ineffective. After irradiation, crRNA system became active due to the RNA linearization. Linear photocleavable blocking oligonucleotides form the duplex with guide RNA and inhibit the DNA cleavage before irradiation. Cyclic photocleavable blocking oligonucleotides linearize upon irradiation and deactivate CRISPR/Cas9 system. Using of azobenzene non-nucleotide linker within the guide RNA structure allows switching the activity of the system by irradiation with UV (365 nm) or visible (505 nm) light. The theophylline-binding RNA aptamer was introduced into guide RNA structure for allosteric regulation of the CRISPR/Cas9 system, making it possible to turn it on or off by changing the concentration of theophylline.

Conclusion: We proposed a set of approaches for design of CRISPR/Cas9 systems which can be regulated by irradiation or by small molecules.

Funding: The study is supported by RSF grant (No. 22-14-00294).

doi 10.18699/bgrs2024-9.3-10

Generation of iPSCs from a patient with the M694V mutation in the mefv gene associated with familial mediterranean fever and their differentiation into macrophages

Zakharyan R.[1, 2]*, Grigor'eva E.V.[3, 4, 5], Karapetyan L.V.[4], Malakhova A.A.[3, 4, 5], Medvedev S.P.[3, 4, 5], Minina J.M.[3], Hayrapetyan V.H.[1], Vardanyan V.S.[6, 7], Zakian S.M.[3, 4, 5], Arakelyan A.[1, 2]

[1] *Russian-Armenian (Slavonic) University, Yerevan, Armenia*

[2] *Institute of Molecular Biology NAS RA, Yerevan, Armenia*

[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[4] *Meshalkin National Medical Research Center, Ministry of Health of the Russian Federation, Novosibirsk, Russia*

[5] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

* *roksana.zakharyan@rau.am*

Key words: Familial Mediterranean fever; macrophages; patient-specific induced pluripotent stem cells; differentiation; MEFV gene

Motivation and Aim: Familial Mediterranean fever (FMF) is a systemic autoinflammatory disorder caused by inherited mutations in the *MEFV* (Mediterranean FeVer) gene, located on chromosome 16 (16p13.3) and encoding the pyrin protein. Despite the existing data on *MEFV* mutations, the exact mechanism of their effect on the development of the pathological processes leading to the spontaneous and re-current autoinflammatory attacks observed in FMF, remains unclear. Induced pluripotent stem cells (iPSCs) are considered an important tool to study the molecular genetic mechanisms of vari-ous diseases due to their ability to differentiate into any cell type, including macrophages, which contribute to the development of FMF.

Methods and Algorithms: In this study, we developed iPSCs from an Armenian patient with FMF carrying the M694V, p.(Met694Val) (c.2080A>G, rs61752717) pathogenic mutation in exon 10 of the *MEFV* gene.

Results: As a result, macrophages expressing CD14 and CD45 surface markers were obtained. We found that the morphology of macrophages de-rived from iPSCs of a patient with the *MEFV* mutation significantly differed from that of macrophages derived from iPSCs of a healthy donor carrying the wild-type MEFV gene.

Conclusion: Using patient-specific iPSCs from FMF patients and the CRISPR/Cas9 genome editing system, it will be possible to generate modified isogenic iPSC lines with the corrected mutation, as well as introduce the mutation into control "healthy" iPSCs in the future. Thus, it will be possible to study, on isogenic lines, the contribution of this mutation to changes not only in the morphology, but also in the functional characteristics of macrophages. Such cell platforms will be valuable for understanding the effects of the mutations on pyrin inflammasome dysfunction in FMF.

Funding: The cell reprogramming and characterization was funded by the Ministry of Science and Higher Education of the Russian Federation, Agreement No. 075-15-2021-1063/10. PBMC isolation and molecular-genetic characterization work was supported by the Higher Science and Education Committee of the Ministry of Science, Education,

Culture and Sports of the Republic of Armenia, in the frames of the research project N 21SCG-1F010 and research grant provided by the Armenian Engineers and Scientists of America (AESA, PI: Dr. Roksana Zakharyan). The immunofluorescent imaging was performed using resources of the Common Facilities Center of Microscopic Analysis of Biological Objects, ICG SB RAS (https://ckp.icgen.ru/ckpmabo/, accessed on 13 March 2024), supported by the Budget project of the Institute of Cytology and Genetics (FWNR-2022-0015).

References

1. Chae J.J., Aksentijevich I., Kastner D.L. Advances in the understanding of familial Mediterranean fever and possibilities for targeted therapy. *Br J Haematol*. 2009;146:467-478
2. Booty M.G., Chae J.J., Masters S.L. et al. Familial Mediterranean fever with a single MEFV mutation: Where is the second hit? *Arthritis Rheum*. 2009;60:1851-1861. doi 10.1002/art.24569
3. Cantarini L., Rigante D., Brizi M.G. et al. Clinical and biochemical landmarks in systemic autoinflammatory diseases. *Ann Med.* 2012;44:664-673. doi 10.3109/07853890.2011.598546

9 Симпозиум «Биомедицина, биоинформатика и системная компьютерная биология»

Symposium “Biomedicine, bioinformatics and systems computational biology”

doi 10.18699/bgrs2024-9.4-01

Молекулярное моделирование механизма противовирусного действия новых конъюгатов 1,7,7-триметилбицикло[2.2.1]гептана и насыщенных N-гетероциклов, связанных 1,2,3-триазольным линкером, в отношении вируса Марбург

Баев Д.С., Соколова А.С., Яровая О.И., Толстикова Т.Г., Салахутдинов Н.Ф.
Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия
** baev@nioch.nsc.ru*

Ключевые слова: монотерпеноиды; филовирусы; ингибиторы; молекулярное моделирование

Мотивация и цель: в продолжение исследований производных монотерпеноидов, проявляющих противовирусное действие в отношении филовирусов, было сделано предположение, что эти молекулы могут ингибировать вирусный гликопротеин вируса Марбург (MARV GP). По данным литературы[1], гомология по первичной последовательности между гликопротеином вируса Эбола (EBOV GP) и MARV GP составляет лишь 27%. Однако, известные ингибиторы EBOV GP, такие как торемифен, способны ингибировать MARV GP, хотя и в более высоких концентрациях[2]. Направленный мутагенез аминокислот сайта связывания EBOV GP, важных для взаимодействия с ингибиторами, приводит к появлению зависимости эффект/концентрация, характерной для ингибирования псевдовирусных частиц, содержащих MARV GP дикого типа. Таким образом, был сделан вывод о том, что на поверхности тримера вирусного гликопротеина филовирусов существует дополнительный сайт связывания ингибиторов, который в случае MARV GP является основным. Такой сайт связывания может находиться в домене HR2 гликопротеинового тримера[3]. HR2 представляет собой преимущественно α-спиральный домен, который соединяет HR1 на N-конце домена GP2 с вирусной мембраной на C-конце. В процессе слияния домены HR1 и HR2 складываются и образуют трансмембранную структуру, состоящую из шести спиралей[4]. HR2-подобные структуры являются весьма консервативными и входят в состав поверхностных гликопротеинов большого числа вирусов разных классов, выполняя одинаковую функцию слияния мембран вирусов и клеток-хозяев. Целью исследования стало изучение теоретической аффинности новых производных к возможному сайту связывания в домене HR2 тримера MARV GP.
Методы и алгоритмы: молекулярный докинг осуществлялся в среде визуализации Schrodinger Maestro с использованием приложений из пакета Schrodinger Small Molecule Drug Discovery Suite 2020-2. Трехмерные структуры производных были получены эмпирически в приложении LigPrep с использованием силового поля OPLS4. Для расчетов применялась рентгеноструктурная модель гликопротеина вируса Марбург, сокристаллизованная с моноклональными антителами[5] (PDB ID 6BP2, разрешение 3,17 Å). Для моделирования возможного механизма связывания с выбранной мишенью выполнялся молекулярный докинг с использованием

протокола IFD[6] (induced fit docking), который использует программы Glide и Prime для предикции положений лигандов в сайте связывания с учетом их влияния на структуру мишени. Область поиска для расчетной функции докинга была выбрана вручную, исходя из положения кармана, образуемого С-концами α-спиралей доменов HR2 тримера MARV GP. Применялся алгоритм повышенной точности докинга XP (extra precision). Докинг проводился в сравнении с результатами IFD, полученными для ингибиторов торемифена и сертралина. Были применены следующие условия: гибкий белок и лиганды, размер области докинга 20 Å, аминокислоты в пределах 5 Å от лиганда учитывались для оптимизации его влияния. Свободную энергию (ΔG) образовавшихся в результате выполнения протокола IFD нековалентных комплексов рассчитывали методом MM/GBSA[7] с использованием Prime. Расчет pKa четвертичных атомов азота пироллидинового, пиперидинового и морфолинового цикла новых производных монотерпеноидов выполнялся с помощью графовой конволюционной нейронной сети веб-сервера MolGpKa[8]. Нековалентные взаимодействия соединений в сайте связывания визуализировались с помощью Schrodinger Maestro.

Результаты: предполагаемый сайт связывания домена HR2 тримера MARV GP представляет собой полость, образуемую тремя идентичными С-концами α-спиралей. Связывание лиганда в этой полости приводит к невозможности конформационных изменений, приводящих к слиянию мембран вируса и лизосомы клетки-хозяина. Центральная часть сайта связывания обладает преимущественно гидрофобным характером, а участки α-спиралей вблизи возможного пространства связывания лигандов насыщены полярными аминокислотами, способными формировать водородные связи с полярными атомами возможных ингибиторов, которые могут проникать в промежутки между α-спиралями.

Из проведенных расчетов pKa четвертичных атомов азота гетероциклов новых конъюгатов следует, что для всех новых соединений при уровне pH, характерном для внутренней среды лизосом, преобладают формы с протонированным четвертичным атомом азота гетероцикла. Появление атома кислорода в составе морфолинового цикла существенно снижает теоретическую аффинность соединений к домену HR2 тримера MARV GP. При этом соединение, обладающее самым длинным линкером, является наиболее активным среди соединений с морфолиновым циклом в результате молекулярного моделирования. Разброс расчетного параметра docking score среди рядов соединений с пиперидиновым и морфолиновым циклом не велик, однако, в ряду соединений с морфолиновым циклом наиболее активным ингибитором в результате расчетово является соединение, также выделяющееся самым длинным линкерным фрагментом.

Рассматривая особенности нековалентных взаимодействий соединений можно сделать вывод о том, что фармакофорная структура вероятных ингибиторов домена HR2 тримера MARV GP, состоящая из объемного гидрофобного ядра и линкерного участка с гидрофильным четвертичным атомом азота хорошо соответствует особенностям сайта связывания. Монотерпеноидное ядро новых соединений располагается в центральной гидрофобной части сайта связывания. Протоны положительно заряженных четвертичных атомов азота гетероциклов новых соединений могут активно участвовать в формировании водородных связей, проникая в гидрофильные борозды между α-спиралями домена HR2. Наблюдаются также водородные связи, возникающие и за счет акцепторных

атомов ингибиторов, как эфирного атома кислорода, так и атома кислорода в составе морфолинового цикла. Не прослеживается существенной разницы в активности соединений, содержащих пятичленный пироллидиновый и шестичленный пиперидиновый циклы.
Выводы: новые конъюгаты теоретически способны взаимодействовать с сайтом связывания, образованным С-концевыми участками α-спиралей домена HR2 тримера MARV GP. Наличие атома кислорода в составе морфолинового цикла новых конъюгатов существенно снижает теоретическую аффинность соединений в отношении сайта связывания домена HR2 MARV GP. Увеличение длины алифатического линкера между триазольным циклом и гетероциклом в структуре соединений, содержащих пиперидиновый и морфолиновый цикл повышает их теоретическую аффинность к рассматриваемому сайту связывания. Эта тенденция нарушается в ряду соединений с пирролидиновым циклом, соединение с самым коротким линкером демонстрирует самое низкое значение docking score в результате расчетов. Результаты молекулярного моделирования были подтверждены изучением ингибирующей активности новых конъюгатов в отношении псевдовирусных частиц, содержащих гликопротеин вируса Марбург. Полученные данные были опубликованы в статье [9].
Финансирование: исследование поддержано Министерством науки и высшего образования РФ (государственное задание № 075-03-2023-642).

Molecular modeling of the mechanism of antiviral action of new conjugates of 1,7,7-trimethylbicyclo[2.2.1]heptane and saturated N-heterocycles linked by a 1,2,3-triazole linker against the Marburg virus

Baev D.S.*, Sokolova A.S., Yarovaya O.I., Tolstikova T.G., Salakhutdinov N.F.
N.N. Vorozhtsov Novosibirsk Institute of Organic chemistry SB RAS, Novosibirsk, Russia
** baev@nioch.nsc.ru*

Key words: monoterpenoids; filoviruses; inhibitors; molecular modeling

Motivation and Aim: in continuation of studies of monoterpenoid derivatives exhibiting antiviral effects against filoviruses, it was suggested that these molecules could inhibit the Marburg virus viral glycoprotein (MARV GP). According to the literature [1], the primary sequence homology between the Ebola virus glycoprotein (EBOV GP) and MARV GP is only 27 %. However, known EBOV GP inhibitors, such as toremifene, are able to inhibit MARV GP, although at higher concentrations [2]. Targeted mutagenesis of amino acids at the EBOV GP binding site that are important for interaction with inhibitors results in an effect/concentration relationship characteristic of the inhibition of pseudoviral particles containing wild-type MARV GP. Thus, it was concluded that on the surface of the filovirus viral glycoprotein trimer there is an additional inhibitor binding site, which in the case of MARV GP is the main one. Such a binding site may be located in the HR2 domain of the glycoprotein trimer [3]. HR2 is a predominantly α-helical domain that connects HR1 at the N terminus of the GP2 domain to the viral membrane at the C terminus. During fusion, the HR1 and HR2 domains fold and form a

transmembrane structure consisting of six helices [4]. HR2-like structures are very conservative and are part of the surface glycoproteins of a large number of viruses of different classes, performing the same function of fusion of the membranes of viruses and host cells. The aim of the research was to study the theoretical affinity of the new derivatives for a possible binding site in the HR2 domain of MARV GP trimer.

Methods and Algorithms: molecular docking was performed in the Schrodinger Maestro visualization environment using applications from the Schrodinger Small Molecule Drug Discovery Suite 2020-2. The 3D structures of the derivatives were obtained empirically in the LigPrep application using the OPLS4 force field. For calculations, we used the X-ray structural model of the Marburg virus glycoprotein, cocrystallized with monoclonal antibodies [5] (PDB ID 6BP2, resolution 3.17 Å). To model the possible mechanism of binding to the selected target, molecular docking was performed using the IFD[6] (induced fit docking) protocol, which uses the Glide and Prime programs to predict the positions of ligands in the binding site, taking into account their influence on the structure of the target. The search region for the estimated docking function was manually selected based on the position of the pocket formed by the C-termini of the α-helices of the HR2 domains of the MARV GP trimer. The XP (extra precision) docking algorithm was used. Docking was carried out in comparison with the IFD results obtained for the inhibitors toremifene and sertraline. The following conditions were applied: flexible protein and ligands, docking region size 20 Å, amino acids within 5 Å of the ligand were considered to optimize its effect. The free energy (ΔG) of non-covalent complexes formed as a result of the IFD protocol was calculated by the MM/GBSA method [7] using Prime. Calculation of pKa of the quaternary nitrogen atoms of the pyrollidine, piperidine and morpholine rings of new monoterpenoid derivatives was performed using the graph convolutional neural network of the MolGpKa web server [8]. Noncovalent interactions of compounds at the binding site were visualized using Schrodinger Maestro.

Results: the putative binding site for the HR2 domain of the MARV GP trimer is a cavity formed by three identical C-termini of α-helices. Binding of the ligand in this cavity prevents the conformational changes leading to the fusion of the membranes of the virus and the host cell lysosome. The central part of the binding site is predominantly hydrophobic, and sections of the α-helices near the possible ligand binding space are saturated with polar amino acids capable of forming hydrogen bonds with the polar atoms of possible inhibitors, which can penetrate into the spaces between the α-helices. From the calculations of pKa of the quaternary nitrogen atoms of heterocycles of new conjugates, it follows that for all new compounds at the pH level characteristic of the internal environment of lysosomes, forms with a protonated quaternary nitrogen atom of the heterocycle predominate. The appearance of an oxygen atom in the morpholine ring significantly reduces the theoretical affinity of compounds for the HR2 domain of the MARV GP trimer. Moreover, the compound with the longest linker is the most active among compounds with a morpholine ring as a result of molecular modeling. The scatter of the calculated docking score parameter among the series of compounds with a piperidine and morpholine ring is not large; however, in the series of compounds with a morpholine ring, the most active inhibitor, as a result of the calculation, is the compound that also has the longest linker fragment.

Considering the features of non-covalent interactions of compounds, we can conclude that the pharmacophore structure of potential inhibitors of the HR2 domain of the MARV GP trimer, consisting of a hydrophobic core and a linker region with a hydrophilic

quaternary nitrogen atom, corresponds well to the characteristics of the binding site. The monoterpenoid core of the new compounds is located in the central hydrophobic part of the binding site. Protons of the positively charged quaternary nitrogen atoms of the heterocycles of the new compounds can actively participate in the formation of hydrogen bonds, penetrating into the hydrophilic grooves between the α-helices of the HR2 domain. Hydrogen bonds are also observed that arise due to the acceptor atoms of the inhibitors, both the ether oxygen atom and the oxygen atom in the morpholine ring. There is no significant difference in the activity of compounds containing a five-membered pyrollidine and six-membered piperidine rings.

Conclusion: the new conjugates are theoretically capable of interacting with the binding site formed by the C-termini of the α-helices of the HR2 domain of the MARV GP trimer. The presence of an oxygen atom in the morpholine ring of the new conjugates significantly reduces the theoretical affinity of the compounds for the binding site of the HR2 domain of MARV GP. Increasing the length of the aliphatic linker between the triazole ring and the heterocycle in the structure of compounds containing a piperidine and morpholine ring increases their theoretical affinity for the binding site in question. This trend is broken in the series of compounds with a pyrrolidine ring; the compound with the shortest linker shows the lowest docking score in the calculations. The results of molecular modeling were confirmed by studies of the inhibitory activity of new conjugates against pseudoviral particles containing MARV GP. Based on the results of the research, an article was published [9].

Funding: the study was supported by the Ministry of Science and Higher Education of the Russian Federation (state assignment No. 075-03-2023-642).

Список литературы/References

1. Ren J., Zhao Y., Fry E.E., Stuart D.I. Target Identification and Mode of Action of Four Chemically Divergent Drugs against Ebolavirus Infection. *J Med Chem.* 2018;61:724-733. doi 10.1021/acs.jmedchem.7b01249
2. Cooper L., Galvan Achi J., Rong L. Comparative analyses of small molecule and antibody inhibition on glycoprotein-mediated entry of Měnglà virus with other filoviruses. *J Med Virol.* 2022;94:3263-3269. doi 10.1002/jmv.27739
3. Schafer A., Xiong R., Cooper L. et al. Evidence for distinct mechanisms of small molecule inhibitors of filovirus entry. *PLoS Pathog.* 2021;17:e1009312. doi 10.1371/JOURNAL.PPAT.1009312
4. Porotto M., Yokoyama C.C., Palermo L.M. et al. Viral Entry Inhibitors Targeted to the Membrane Site of Action. *J Virol* 2010;84:6760-6768. doi 10.1128/jvi.00135-10
5. King L.B., Fusco M.L., Flyak A.I. et al. The Marburgvirus-Neutralizing Human Monoclonal Antibody MR191 Targets a Conserved Site to Block Virus Receptor Binding. *Cell Host Microbe.* 2018;23:101-109.e4. doi 10.1016/j.chom.2017.12.003
6. Sherman W., Day T., Jacobson M.P., Friesner R.A., Farid R. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem.* 2006;49:534-553. doi 10.1021/jm050540c
7. Genheden S., Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov.* 2015;10:449-461
8. Pan X., Wang H., Li C., Zhang J.Z.H., Ji C. MolGpka: A Web Server for Small Molecule pKa Prediction Using a Graph-Convolutional Neural Network. *J Chem Inf Model.* 2021;61:3159-3165. doi 10.1021/acs.jcim.1c00075
9. Sokolova A.S., Yarovaya O.I., Artyushin O.I. et al. Design, synthesis and antiviral evaluation of novel conjugates of the 1,7,7-trimethylbicyclo[2.2.1]heptane scaffold and saturated N-heterocycles via 1,2,3-triazole linker. *Arch Pharm*. 2024;357:2300549. doi 10.1002/ardp.202300549

doi 10.18699/bgrs2024-9.4-02

Производные дигидробетулоновой кислоты как потенциальные двойные агонисты PPAR-α и PPAR-γ рецепторов

Борисов С.*, Блохин М., Хвостов М., Фоменко В., Лузина О., Салахутдинов Н.
Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия
** sergalborisov@mail.ru*

Ключевые слова: сахарный диабет; гипогликемическая активность; ОГТТ, глитазары; терпеноиды, дигидробетулоновая кислота

Мотивация и цель: Сахарный диабет второго типа (СД2), характеризуемый резистентностью к инсулину и сопутствующей гипергликемией, является в наше время глобальной проблемой здравоохранения, так как данное заболевание охватывает уже более 10 % мирового населения, и эта цифра постоянно растет [1]. В связи с такой тенденцией в последние десятилетия идет постоянный поиск биологических мишеней для разработки новых препаратов, способных снизить выраженность основных симптомов СД2. К сожалению, существующие подходы для лечения СД2 пока что далеки от оптимальных вследствие низкой эффективности используемых препаратов и обширного списка их побочных эффектов. Перспективными мишенями в этом направлении является подсемейство ядерных рецепторов – рецепторов, активируемых пероксисомными пролифераторами (peroxisome proliferator-activated receptor, PPAR), активация которых может нормализовать метаболические дисфункции и уменьшить сердечно-сосудистые факторы риска, связанные с СД2. Основными классами PPAR являются PPAR-α и PPAR-γ, агонисты которых могут корректировать дислипидемию и снижать резистентность тканей к инсулину соответственно. Наибольший интерес вызывают двойные лиганды PPAR α и γ, названные глитазарами и сочетающие в одной молекуле гиполипидемические и гипогликемические свойства α и γ-агонистов. При этом, почти все разработанные на настоящий момент глитазары имели побочные эффекты, которые варьировались в зависимости от структурного фрагмента, связанного с их консервативной частью, представленной в абсолютном большинстве случаев дезаминированным L-тирозином. В связи с этим введение природных фармакофоров может привести к существенному улучшению свойств потенциальных лекарственных средств, поскольку природные соединения имеют высокий профиль безопасности, выработанное в процессе эволюции структурное сродство ко многим биологическим рецепторам. Целью данного исследования была оценка гипогликемического и гиполипидемического действия пяти новых потенциальных глитазаров с фрагментом дигидробетулоновой кислоты, отделенным от консервативной части линкерами различной длинны (от 2 до 6 атомов углерода).
Методы и алгоритмы: В опытах участвовали самцы мышей линии C57BL/6 $A^{y/a}$ с ожирением, которые за 30 дней до начала эксперимента и в течение всего эксперимента находились на диете с высоким содержанием жиров и углеводов.

Животным ежедневно в течение 28 дней (4 недели) вводили внутрижелудочно исследуемые соединения в дозе 30 мг/кг, в качестве препаратов сравнения использовали тезаглитазар в дозе 30 мг/кг и метформин в дозе 250 мг/кг, а животным контрольной группы вводили растворитель. Также отдельно, в качестве интактной контрольной группы использовали самцов мышей линии C57BL/6. На 15-й и 30-й день эксперимента был проведен оральный глюкозотолерантный тест (ОГТТ) для оценки потенциального гипогликемического эффекта изучаемых соединений. Дополнительно, еженедельно фиксировалась масса животных и количество потребленного корма.

Результаты: По результатам измерения массы тела мышей было установлено что в группе тезаглитазара и потенциального глитазара с наименьшей длиной линкера (2 атома) существенно снизить этот показатель в течение эксперимента (на 32 и 19 % соответственно). У мышей в остальных группах масса сохранялась примерно на одном уровне и соответствовала состоянию выраженного ожирения. Как было видно по данным, полученным в ходе ОГТТ на 15й день эксперимента, резистентность к инсулину развилась у всех мышей линии C57BL/6 $A^{y/a}$, что является основным симптомом СД2. При этом следует отметить, что все исследуемые соединения, так же, как и препараты сравнения тезаглитазар и метформин, проявили достоверное по отношению к контрольной группе гипогликемическое действие. Результаты ОГТТ, проведенного на 30-й день, в целом соответствовали первому ОГТТ, однако отсутствовал эффект в группе метформина, что объяснялось тем, что вещества вводили последний раз за два дня до проведения теста (его эффект длится не более 12 ч). Кроме того, после подсчета площади под гликемической кривой было обнаружено, что наиболее выраженным гипогликемическим действием обладали тезаглитазар и соединение с самым коротким линкером, что согласовывалось с данными измерения массы.

Выводы: Исходя из полученных в ходе эксперимента данных, было обнаружено, что все потенциальные глитазары с фрагментом дигидробетулоновой кислоты в определенной степени проявляют гипогликемическое действие на мышах с симптомами, характерными для СД2, однако наиболее активным из них было соединение с наиболее коротким линкером (2 атома углерода). При этом данное вещество, в отличие от остальных, подобно тезаглитазару, способствовало существенному снижению массы тела животных. Таким образом, можно сделать вывод, что наличие тритерпенового фрагмента, соединенного с консервативной частью молекулы глитазара линкером с двумя атомами углерода, способствует наиболее выраженному гипогликемическому эффекту и, вероятно, связанному с этим повышению чувствительности к инсулину.

Финансирование: Исследование выполнено при поддержке гранта РНФ «Разработка лекарственных агентов нового типа для лечения заболеваний сердечно-сосудистой системы, связанных с нарушениями метаболизма липидов и глюкозы» (№ 24-25-00120).

Dihydrobetulonic acid derivatives as potential dual PPAR-α and PPAR-γ agonists

Borisov S.*, Blokhin M., Khvostov M., Fomenko V., Luzina O., Salakhutdinov N.
N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry SB RAS, Novosibirsk, Russia
** sergalborisov@mail.ru*

Key words: diabetes mellitus; hypoglycemic activity; OGTT, glitazars; terpenoids, dihydrobetulonic acid

Motivation and Aim: Type 2 diabetes mellitus (DM2), characterized by insulin resistance and subsequent hyperglycemia is nowadays a global healthcare issue affecting more than 10 % of the world population with this number continuously growing [1]. Due to this trend, in the last decades, there is a constant search for new biological targets for the development of drugs that would alleviate the main symptoms of DM2. Unfortunately, the existing DM2 treatment approaches are still far from optimal, which is a consequence of the low efficiency of the drugs used and their plentiful side effects. A subfamily of nuclear receptors – peroxisome proliferator-activated receptors (PPAR) is a promising target in this research direction, because their activation is able to normalize the possible metabolic dysfunctions and reduce the cardiovascular risks associated with DM2. The main classes of PPARs are PPAR-α and PPAR-γ, which agonists can reverse dyslipidemia and reduce insulin resistance respectively [2]. Lately dual ligands of PPAR α and γ, called glitazars, are becoming of most interest due to their ability to combine the hypolipidemic and hypoglycemic properties of PPAR α and γ agonists in one molecule. Yet, to this day, almost all of the developed glitazars had adverse effects, which varied based on the structural fragment, connected to the conservative part of the molecule, usually represented by deaminated L-tyrosine [3]. In this regard, the introduction of natural pharmacophores, terpenoids in particular, may lead to a significant improve in the properties of the drugs, due to them having a high safety profile, an evolutionary developed structural affinity to a lot of biological receptors and a wide range of their own biological activity [4]. The aim of this study was to evaluate the hypoglycemic and hypolipidemic activity of five new potential glitazars with a dihydrobetilonic acid fragment, separated from the conservative part by linkers of different length (from 2 up to 6 carbon atoms).
Methods and Algorithms: In our experiments obese male C57BL/6 $A^{y/a}$ mice were used, which 30 days prior and during the experiment were kept on a high-fat and high-carbohydrate diet. Animals daily for 28 days were injected intragastrically with the test compounds at a dose of 30 mg/kg, as reference drugs tesaglitazar at a dose of 30 mg/kg and metformin at a dose of 250 mg/kg were used, while animals of the control group were treated with the solvent only. Apart from that, male C57BL/6 mice were used as an intact control group. On the 15^{th} and 30^{th} day of the experiment, oral glucose tolerance tests were carried out in order to assess the potential hypoglycemic effect of the tested compounds. Additionally the animals' weight and the amount of chow consumed were recorded weekly.
Results: According to the results of the animals' weight measurements in the groups of tesaglitazar and the potential glitazar with the shortest linker (2 carbon atoms), a significant reduction of this parameter (32 and 19 % respectively) during the experiment was noticed. The mean mice weight value in all other groups remained approximately at

the same level and corresponded to the state of severe obesity. Resistance to insulin, which is the main symptom of DM2, developed in all C57BL/6 $A^{y/a}$ mice, as was evident from the obtained data of the OGTT on the 15th experiment day. That said it is worth noting that all tasted compounds as well as both reference drugs, tesaglitazar and metformin, showed a statistically significant hypoglycemic effect compared to the control group. The second OGTT (30th day) results were generally in line with the first OGTT data, except for the effect absence in the metformin group, which can be explained by the fact that substances were last administered 2 days before the test (metformin's action lasts no longer than 12 hours). Moreover after the area under the glycemic curve calculation tesaglitasar and the test compound with the shortest linker were found to possess the most pronounced hypoglycemic activity among all other substances under investigation, which was on par with the weight measurement data.

Conclusion: According to the data obtained during the experiment it was found that all potential glitazars with the dihydrobetulonic acid fragment to a certain extent exhibit hypoglycemic effects in mice with DM2 symptoms, yet the most active of them was the compound with the shortest linker (2 carbon atoms). At the same time, this compound, unlike the other dihydrobetulonic acid derivatives with longer linkers and similar to tesaglitazar, contributed to a significant animal weight loss. Thus, it could be concluded that the most pronounced hypoglycemic effect, which is most probably related to increased insulin sensitivity is achieved by connecting the conservative part of the glitazar molecule to a dihydrobetulonic acid fragment through a linker, consisting of 2 carbon atoms.

Funding: The study is supported by the RSF grant "Development of new type of drug agents for the treatment of cardiovascular diseases associated with lipid and glucose metabolism disorders" (No. 24-25-00120).

Список литературы/References

1. Ruze R. et al. Obesity and type 2 diabetes mellitus: connections in epidemiology, pathogenesis, and treatments. *Front Endocrinol*. 2023;14:1161521. doi 10.3389/fendo.2023.1161521
2. Mirza A.Z., Althagafi I.I., Shamshad H. Role of PPAR receptor in different diseases and their ligands: Physiological importance and clinical implications. *Eur J Med Chem*. 2019;15(166):502-513. doi 10.1016/j.ejmech.2019.01.067
3. Cheng H.S. et al. Exploration and development of PPAR modulators in health and disease: an update of clinical evidence. *Int J Mol Sci*. 2019;20(20):5055. doi 10.3390/ijms20205055
4. Chukwujekwu J. et al. Alpha-glucosidase inhibitory and antiplasmodial properties of terpenoids from the leaves of Buddle jasaligna Willd. *J Enzyme Inhib Med Chem*. 2015;31:63-66. doi 10.3109/14756366.2014.1003927

doi 10.18699/bgrs2024-9.4-03

Ингаляционная доставка нано- и субмикронных частиц антибактериальных агентов

Валиулин С.[1]*, Аньков С.[1, 2], Толстикова Т.[1, 2], Бакланов А.[1], Дульцева Г.[1], Онищук А.[1]

[1] *Институт химической кинетики и горения им. В.В. Воеводского СО РАН, Новосибирск, Россия*

[2] *Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия*

** valiulin@kinetics.nsc.ru*

Ключевые слова: нано- и субмикронный частицы; аэрозоль; антибактериальные вещества; ингаляция

Мотивация и цель: Серьезным вызовом для современной системы здравоохранения является развитие бактериальной устойчивости к большинству известных антибактериальных средств. По данным Всемирной организации здравоохранения [1] бактериальные инфекции и опосредованные ими заболевания входят в топ-10 причин смерти в мире. Смертность от устойчивых штаммов составляет порядка 700 тыс. человек в год и продолжает расти, что к 2050 году может привести к десяткам миллионов смертей по всему миру. Для преодоления этой проблемы проводят поиск новых типов антибактериальных агентов и разработку способов повышения эффективности уже существующих лекарственных средств. Так, адресная доставка с созданием высокой локальной концентрации действующего вещества может потенциально позволить преодолеть устойчивость и в то же время снизить общую лекарственную нагрузку на организм. При поражении дыхательной системы таким методом адресной доставки является ингаляция, когда лекарственное средство в виде мелкодисперсных частиц движется по каналам дыхательной системы и оседая локально в ней оказывает свое терапевтическое действие. Однако такой способ требует технических средств, позволяющих осуществлять генерацию частиц заданного размерного диапазона для их осаждения в определенных пораженных участках дыхательной системы и точного контроля доставляемой таким способом дозы лекарственного средства.

Целью настоящей работы являлась разработка метода генерации аэрозольных частиц антибиотиков цефалоспоринового ряда в нано- и субмикронном диапазоне размеров и исследование фармакокинетики и специфической активности после их ингаляционной доставки.

Методы и алгоритмы: Исследование было проведено на антибактериальных препаратах цефазолин, цефуроксим, цефтриаксон и цефепим. Генерация аэрозоля осуществлялась методом ультразвукового распыления водного раствора лекарственных веществ. Принцип работы блока генератора аэрозоля заключается в следующем: в дно цилиндрической стеклянной кюветы, объемом 200 см^3, встроен пьезоэлемент, а раствор исследуемого вещества (≈20 см^3) заливается в кювету через боковой штуцер. За счет колебаний пьезоэлемента происходит образование аэрозоля в виде капель раствора. Этот аэрозоль увлекается потоком фильтрованного воздуха (Q_1 = 0,3 л/мин), который подается через боковой штуцер

кюветы. Поток аэрозоля из кюветы поступает в вертикальный холодильник, охлаждаемый холодной водой (≈10 °С). Здесь происходит конденсация на стенки паров воды, которые были захвачены потоком воздуха и частичное испарение растворителя из капель аэрозоля. На выходе из холодильника поток аэрозоля Q_1 смешивается с потоком сухого воздуха (Q_2 = 2,8 л/мин) и сформировавшийся, новый аэрозольный поток (Q_1 + Q_2) движется из блока генерации аэрозоля в осушитель, в котором происходит окончательное испарение растворителя и формирование сухих (твердых) аэрозольных частиц. Размер частиц и концентрацию на выходе из генератора определяли с помощью диффузионного аэрозольного спектрометра ДСА-М и оптического спектрометра аэрозоля ОСА, которые вместе позволяют измерять счетную концентрацию частиц в диапазоне от 10 до $5 \cdot 10^5$ см$^{-3}$ и распределение по размерам в диапазоне от 0,003–10 мкм [2].

Ингаляционную доставку генерируемого аэрозоля лабораторным мышам проводили в ингаляционной камере «nose-only», оснащенной двенадцатью радиально расположенными портами (2 яруса по 6 портов), куда вставляются специальные стеклянные объемы, внутри которых животное располагается таким образом, чтобы только нос находился в области движения потока аэрозоля.

Для сравнения с ингаляционной формой доставки, выполняли исследование фармакокинетики после внутривенного и внутрибрюшного введения.

Количественное определение лекарственного вещества в сыворотке крови и легких осуществляли методом ВЭЖХ на жидкостном хроматографе «Милихром А-02».

Проводилось исследование специфической активности аэрозольной формы в сравнении с внутрибрюшинным и внутривенным введением антибактериальных веществ на модели бактериального сепсиса у мышей вызванного путем внутрибрюшинного введения суспензии бактерий в количестве $5 \cdot 10^5$ КОЕ на животное. Использовались два музейных штамма бактерий: *Klebsiella pneumoniae* 82 и *Staphylococcus aureus* ATCC 25 953. Обе культуры бактерий были получены из коллекции ФГБУН СФНЦА РАН.

Результаты: Методом ультразвукового распыления водного раствора был получен аэрозоль антибиотиков цефалоспоринового ряда. Размер аэрозольных частиц зависел от концентрации и природы вещества в распыляемом растворе и составлял, в случае:

цефазолина от 80 до 1100 нм, при концентрации раствора от 0.25 до 14 %;
цефуроксима от 250 до 550 нм, при концентрации раствора от 1 до 12 %;
цефтриаксона от 300 до 900 нм и цефепима от 450 до 1300 нм, при концентрации раствора от 0,5 до 14 %.

При этом, концентрация аэрозольных частиц составляла $(2 \pm 0.5) \cdot 10^5$ см$^{-3}$ и практически не зависела от концентрации вещества в растворе. Это обусловлено тем, что под действием ультразвуковых колебаний (при постоянной частоте) происходила генерация примерно одинакового количества жидких капель, содержащих растворенное вещество.

Фармакокинетическое исследование показало, что содержание цефалоспоринов в легких после их ингаляции на порядок выше по сравнению с инъекционным введением. Биодоступность цефалоспоринов при их аэрозольной доставке составила примерно 80 ± 5 %, что сопоставимо с внутрибрюшинным введением.

Аэрозольная форма доставки показала хороший лечебный антибактериальный эффект. Животные, подвергнутые ингаляционному лечению, оставались живы

весь период наблюдения – 9 суток. У всех инфицированных животных увеличилось количество нейтрофилов и уменьшилось количество лимфоцитов по сравнению с интактной группой, что является результатом воспаления. Однако у мышей, получавших ингаляцию, отклонение от интактных животных было в 1,5–2 раза меньше, чем у нелеченых мышей. Аэрозольная форма доставки продемонстрировала хороший лечебный антибактериальный эффект, сопоставимый по эффективности с инъекционным введением. Через 8 ч после инфицирования, у животных, получавших лечение ингаляционно, количество КОЕ в крови оказалась на два порядка меньше, чем у животных, группы отрицательного контроля.

Выводы: Для проведения исследований, связанных с ингаляционной доставкой антибиотиков, разработан и собран экспериментальный стенд, позволяющий выполнять контролируемую генерацию аэрозоля и его ингаляционную доставку.

Был получен аэрозоль антибиотиков цефалоспоринового ряда с размером частиц от 80 до 1100 нм и средней счетной концентрацией $(2 \pm 0.5) \cdot 10^5$ см$^{-3}$.

Фармакокинетическое исследование показало, что содержание цефалоспаринов в легких после их ингаляции на порядок выше по сравнению с инъекционным введением, а их биодоступность при аэрозольной доставке составляет 80 ± 5 % и сопоставима с внутрибрюшинным введением.

Исследование специфического действия антибиотиков цефалоспоринового ряда при их ингаляционном введении на модели бактериального сепсиса продемонстрировало хороший лечебный эффект сопоставимый по эффективности с инъекционным введением.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда (№ 19-73-10143, https://rscf.ru/project/19-73-10143/).

Inhalation delivery of nano- and submicron particles of antibacterial agents

Valiulin S.[1]*, An'kov S.[1, 2], Tolstikova T.[1, 2], Baklanov A.[1], Dultseva G.[1], Onischuk A.[1]

[1] *V.V. Voevodsky Institute of Chemical Kinetics and Combustion, SB RAS, Novosibirsk, Russia*

[2] *N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia*

* *valiulin@kinetics.nsc.ru*

Key words: nano- and submicron particles; aerosol; antibacterial substances; inhalation

Motivation and Aim: A significant challenge for modern public health system is emergence of bacterial resistance against most of the known antibacterial agents. According to the data of the World Health Organization [1], bacterial infections and related diseases are among the top 10 causes of death in the world. Mortality caused by resistant strains accounts for about 700 thousand cases per year and continues to rise, which can lead to tens of millions deaths over the world by 2050. To overcome this problem, search for new types of antibacterial agents and development of the methods to enhance the efficiency of already existing pharmaceuticals are involved. For instance, targeted delivery ensuring the high local concentration of active agents may potentially overcome drug resistance and at the same time reduce the total drug burden on the organism. In the case if respiratory system is affected, a proper method of targeted delivery is inhalation: a pharmaceutical agent in the form of fine particles travels along

the airways of respiratory system, gets locally deposited there and thus causes therapeutic action. However, this method requires adequate technical equipment to generate the particles within the necessary size range to provide their deposition in specific affected sites of the respiratory system and to enable precise control of the delivered drug dose.

The goal of the present work was to develop a method for generating the aerosol particles of cephalosporin antibiotics in the nano- and submicron size range, and to study pharmacokinetics and specific activity after the inhalation delivery of these particles.

Methods and Algorithms: The studies were carried out with antibacterial preparations: cefazolin, cefuroxime, ceftriaxone, and cefepime. The aerosol was generated through ultrasonic spraying of the aqueous solutions of pharmaceutical substances. The operation principle of the aerosol generating unit is as follows: a piezoelectric element is built in the bottom of a cylindrical glass cell 200 cm^3 in volume, and the solution of the substance under investigation (≈20 cm^3) is poured into the cell through a side fitting. The formation of aerosol as the droplets of solution proceeds due to the oscillations of piezoelectric element. This aerosol is carried with the flow of filtered air (Q_1 = 0.3 L/min), supplied through the side fitting of the cell. The aerosol flow from the cell inters a vertical condenser cooled with water (≈10 °C). Here, water vapor captured in the air flow is condensed on the walls, and the solvent is partially evaporated from aerosol droplets. At the outlet of the condenser, the aerosol flow Q_1 is mixed with the dry air flow (Q_2 = 2.8 L/min), and thus generated new aerosol flow (Q_1 + Q_2) passes from the aerosol-generating unit into the dryer, where complete evaporation of the solvent takes place, and dry (solid) aerosol particles are formed. Particle size and concentration at the outlet of the generator were determined with the help of diffusion aerosol spectrometer DSA-M and optical spectrometer of aerosol OSA, which together allow measurement of the number concentration of particles within the range from 10 to $5 \cdot 10^5$ cm^{-3} and size distribution within the range 0.003 – 10 μm [2].

Inhalation delivery of the generated aerosol to laboratory mice was performed in a Nose-Only inhalation chamber equipped with twelve radially located ports (2 storeys with 6 ports in each) to which special glass vessels are mounted. The animals are accommodated in these vessels so that only the nose is within the area of aerosol flow passing through.

To compare with the inhalation delivery, the pharmacokinetic studies were also carried out after intravenous and intraperitoneal introduction.

The quantitative determination of pharmaceutical substances in blood serum and in the lungs was carried out by means of HPLC using Milikhrom A-02 liquid chromatograph.

Specific activity of the aerosol form was studied in comparison with intraperitoneal and intravenous introduction of the antibacterial agents using the model of bacterial sepsis in mice, caused by the intraperitoneal introduction of bacterial suspensions in the amount of $5 \cdot 10^5$ CFU per animal. Two archival bacterial strains were used: *Klebsiella pneumoniae* 82 and *Staphylococcus aureus* ATCC 25 953. Both bacterial cultures were obtained from the collection of the Siberian Federal Research Center of Agrobiotechnologies of the Russian Academy of Sciences.

Results: The aerosol of cephalosporin antibiotics was prepared by ultrasonic spraying of the aqueous solutions. The size of aerosol particles depended on the concentration and nature of the substance in the solution to be sprayed and was determined to be in the case of:

cefazolin, from 80 to 1100 nm, for solution concentration from 0.25 to 14 %; cefuroxime, from 250 to 550 nm, for solution concentration from 1 to 12 %; ceftriaxone, from 300 to 900 nm, and cefepime, from 450 to 1300 nm, for solution concentration from 0.5 to 14 %.

In all these cases, the concentration of aerosol particles was $(2 \pm 0.5)\cdot 10^5$ cm^{-3} and remained practically independent on the concentration of the substance in solution. His is explained by the fact that the action of ultrasonic oscillations (at the constant frequency) generated approximately the same number of liquid droplets containing the dissolved substances.

Pharmacokinetic studies showed that the content of cephalosporins in lungs after inhalation was an order of magnitude higher than after injection. Bioavailability of cephalosporins after aerosol delivery was approximately equal to 80 ± 5 %, which was comparable with the result obtained after intraperitoneal introduction.

The aerosol delivery exhibited good antibacterial effect. The animals subjected to inhalation treatment were alive during the entire time of observation for 9 days. The number of neutrophils increased in all infected animals, while the number of lymphocytes decreased in comparison with the animals of intact group, which was the result of inflammation. However, the difference between the inhalation-treated mice and intact animals was 1.5–2 times smaller than that for non-treated infected mice. The aerosol form demonstrated good curative antibacterial effect, comparable in efficiency with that of injection. After the introduction of infection, 8 hours later, the number of CFU in the blood of inhalation-treated animals was determined to be two orders of magnitude smaller than in the animals of the negative reference group.

Conclusion: To carry out the studies related to the inhalation delivery of antibiotics, we developed and built the experimental test bench allowing controllable aerosol generation and its delivery by inhalation.

The aerosols of cephalosporin antibiotics with particle size 80 to 1100 nm and mean number concentration $(2 \pm 0.5)\cdot 105$ cm^{-3} was obtained.

Pharmacokinetic studies showed that the content of cephalosporins in lungs after their inhalation was an order of magnitude higher than after introduction through injection, and bioavailability of the drugs after inhalation delivery was 80 ± 5%, which was quite comparable with that after intraperitoneal introduction.

Investigation of the specific action of cephalosporin antibiotics after their inhalation delivery in the model of bacterial sepsis demonstrated a good therapeutic effect, comparable with that after injection.

Funding: The study is supported by the Russian Science Foundation (No. 19-73-10143, https://rscf.ru/project/19-73-10143/).

Список литературы/References

1. The top 10 causes of death. Web site: https://www.who.int/ru/news-room/fact-sheets/detail/the-top-10-causes-of-death
2. Valiulin S.V., Onischuk A.A., Baklanov A.M. et al. An integrated aerosol setup for therapeutics and toxicological testing: Generation techniques and measurement instrumentation. *Measurement*. 2021;181:109659

doi 10.18699/bgrs2024-9.4-04

Влияние ингибиторов тирозиновых фосфатаз на гормональные показатели и инсулиновую чувствительность у крыс с диета-индуцированным диабетом 2 типа

Деркач К.*, Зорина И., Шпаков А.
Институт эволюционной физиологии и биохимии им. И.М. Сеченова, Российская академия наук, Санкт-Петербург, Россия
** derkatch_k@list.ru*

Ключевые слова: протеинфосфотирозинфосфатаза 1B; Т-клеточная протеинфосфотирозинфосфатаза; диабет 2 типа; ингибитор; инсулиновый рецептор; лептин; инсулиновая резистентность

Мотивация и цель: Ключевыми патогенетическими факторами при сахарном диабете 2 типа (СД2), самом распространенном эндокринном заболевании в мире, приводящими к множеству осложнений, являются инсулиновая резистентность (ИР), нарушенная толерантность к глюкозе и дислипидемия. Разработка фармакологических подходов, направленных на нормализацию чувствительности к инсулину, глюкозного и липидного гомеостаза, наряду с диетой и физическими нагрузками, является магистральным направлением в лечении СД2 и его осложнений. Одной из причин ИР является повышение активности тирозиновых фосфатаз в ответ на стимуляцию клеток инсулином, которые дефосфорилируют и инактивируют инсулиновый рецептор и сопряженные с ним инсулин-рецепторные субстраты-1 и -2 (IRS1/2), ключевые компоненты инсулинового сигналинга [1, 2]. Среди фосфатаз наиболее важны протеинфосфотирозинфосфатаза 1B (PTP1B) и Т-клеточная протеинфосфотирозинфосфатаза (TCPTP), которые функционируют как негативные регуляторы инсулинового сигналинга. Они также подавляют сигнальные пути лептина, дефосфорилируя лептиновый рецептор и IRS1/2 и вызывая резистентность к лептину [3]. Вследствие этого одним из путей нормализации инсулиновой и лептиновой чувствительности при метаболических расстройствах является применение селективных ингибиторов фосфатаз, в первую очередь PTP1B [4]. Нами разработаны соединения PI4 и Comp3, ингибиторы фосфатаз PTP1B и TCPTP, созданные на основе структуры 4-оксо-1,4-дигидроциннолина, причем показано, что наибольшей эффективностью в отношении улучшения метаболического статуса и снижения ИР характеризовались ингибиторы с двойной специфичностью, ингибирующие как PTP1B, так и TCPTP [5, 6]. Однако их эффективность в отношении коррекции ИР и лептиновой резистентности при СД2 не была изучена, а также не было исследовано их влияние на показатели тиреоидной оси, функции которой нарушены при СД2. Основываясь на этом, целью работы было изучить влияние PI4 и Comp3 на метаболические и гормональные показатели, в том числе на уровни тиреотропного гормона (ТТГ) и тиреоидных гормонов, у самцов крыс с диета-индуцированным СД2.

Материалы и методы: В опытах использовали самцов крыс Wistar, эксперименты осуществляли в соответствии с требованиями Этического комитета ИЭФБ РАН (протокол № 4-2/2023 от 25 апреля 2023 г.). Животных содержали в стандартных условиях вивария в пластиковых контейнерах по 6 крыс в каждом, со свободным доступом к пище и воде. Синтез этил-3-(гидроксиметил)-4-оксо-1,4-дигидроциннолин-6-карбоксилата (PI4) и 6-хлор-3-(гидроксиметил)-циннолин-4(1H)-она (Comp3) осуществляли с помощью реакций Соногаширы и Рихтера, позволяющих сконструировать циннолиновый каркас. Целевые продукты очищали с помощью адсорбционной хроматографии (чистота не менее 98 %), структуру соединений доказывали с использованием масс-спектрометрии высокого разрешения на спектрометре micrOTOF («Bruker», Германия). СД2 вызывали высокожировой диетой продолжительностью 16 недель, на 10-й неделе обрабатывая их стрептозотоцином (20 мг/кг) для декомпенсации функций панкреатических β-клеток. За 2 недели до обработки животных ингибиторами фосфатаз в диабетических группах отбирали животных с выраженными признаками СД2 и ожирения – повышением массы тела в сравнении с контролем не менее чем на 10 % и с нарушенной толерантностью к глюкозе в глюкозотолерантном тесте (ГТТ), что выражалось в ее уровне через 120 мин после глюкозной нагрузки не ниже 7.5 мМ. Затем формировали 4 группы крыс (по 6 животных в каждой) – контроль (C), СД2 без обработки (D), СД2 с 7-дневной обработкой PI4 (Dpi4) или Comp3 (Dcomp3) в суточной дозе 10 мг/кг (раствор в ДМСО). В 7-й день забирали кровь для измерения уровней ТТГ и тиреоидных гормонов. Через сутки после окончания лечения проводили повторный ГТТ, оценивая уровни глюкозы, инсулина и лептина в крови дискретно, на протяжении 120 мин после глюкозной нагрузки. Забор крови осуществляли при местной анестезии (лидокаин, 4 мг/кг). Уровни глюкозы измеряли тест-полосками «One Touch Ultra» (США), уровни инсулина, лептина и ТТГ – с помощью «ELISA for Insulin, Rat» («Cloud-Clone Corp.», США), «ELISA for Leptin, Rat» («Cloud-Clone Corp.», США) и «ELISA for TSH, Rat» («Cusabio Biotech Co., Ltd.», Китай), уровни свободного и общего тироксина (fT4, tT4) и свободного и общего трийодтиронина (fT3, tT3) – с помощью ИФА-наборов фирмы «Иммунотех» (Россия). После этого крыс наркотизировали (хлоралгидрат, 400 мг/кг), декапитировали, оценивали массу тела и жировой ткани. Для статистического анализа использовали пакет программ «Microsoft Office Excel 2007» (США). Результаты представляли, как $M \pm SD$. Нормальность распределения оценивали критерием Шапиро-Уилка. Для сравнения двух выборок с нормальным распределением использовали t-критерий Стьюдента, для сравнения большего числа выборок – дисперсионный анализ с поправкой Бонферрони. Достоверными считали различия при $p < 0.05$.

Результаты: У крыс с СД2 были значимо повышены масса тела и жировой ткани, уровни глюкозы и значение $AUC_{0\text{-}120}$, представляющее собой интегрированную площадь под глюкозной концентрационной кривой в ГТТ, а также базальный уровень лептина и стимулированные глюкозой уровни лептина и инсулина (табл.). Обработка диабетических крыс в течение недели обоими ингибиторами фосфатаз приводила к снижению массы тела и жира, причем в случае PI4 эти показатели не отличались от таковых в контроле, а также к улучшению толерантности к глюкозе и нормализации базовых и стимулированных глюкозой уровней инсулина и лептина. Более чувствительными к обработке ингибиторами фосфатаз были уровни инсулина, причем влияние PI4 на эти показатели было более выраженным

в сравнении с Comp3 (табл. 1). Тем самым ингибиторы PTP1B и TCPTP, характеризующиеся двойной специфичностью, были эффективны для нормализации метаболических и гормональных показателей, что согласуется как с нашими ранними результатами по их применению для коррекции ожирения [5, 6], так и с данными других авторов о восстанавливающем эффекте различных ингибиторов фосфатазы PTP1B на чувствительность тканей к инсулину и лептину, хотя в этом случае эффекты таких ингибиторов при длительном лечении затухали [7]. Это может быть обусловлено тем, что подавление активности PTP1B приводило к компенсаторному усилению экспрессии и активности TCPTP, способной в некоторой степени заменять PTP1B, как негативный регулятор инсулинового и лептинового сигналинга [6].

Таблица 1. Влияние PI4 и Comp3, наделенных двойной специфичностью ингибиторов фосфатаз PTP1B и TCPTP, на массу тела и жировой ткани, уровни глюкозы и гормональный профиль у крыс с диета-индуцированным СД2

Показатель	C	D	Dpi4	Dcomp3
Масса тела, г	345±9	430±11*	381±15#	393±10*#
Масса АЖ, г	3.9±0.5	14.4±1.7*	7.2±0.8*#	8.1±0.7*#
Масса ЭЖ, г	3.7±0.4	7.9±1.2	5.5±0.9	6.4±0.8*
Глюкоза-0, мМ	4.3±0.2	6.1±0.3*	4.0±0.2#	4.6±0.8
Глюкоза-120, мМ	5.1±0.3	9.5±1.3*	5.2±0.6#	6.2±1.4
$AUC_{0\text{-}120}$, усл.ед.	1109±46	1895±65*	1239±57#	1387±48*#
Инсулин-0, нг/мл	0.54±0.10	0.71±0.18	0.47±0.11	0.69±0.11
Инсулин-120, нг/мл	0.68±0.09	1.40±0.15*	0.77±0.10#	0.83±0.11#
Лептин-0, нг/мл	0.95±0.12	2.87±0.43*	1.77±0.18*#	1.79±0.22*#
Лептин-120, нг/мл	1.22±0.38	4.51±0.37*	2.05±0.39#	2.90±0.18*#
ТТГ, мкЕД/мл	0.56±0.09	0.71±0.12	0.50±0.05	0.58±0.07
fT4, пМ	28.8±1.3	22.1±0.8*	26.9±1.1#	25.1±1.9
tT4, нМ	76.2±6.9	69.1±7.0	73.8±4.6	72.2±3.3
fT3, пМ	2.5±0.3	2.1±0.3	2.6±0.3	2.4±0.2
tT3, нМ	3.4±0.2	2.8±0.2*	3.4±0.1#	3.2±0.3

Примечание. АЖ и ЭЖ – абдоминальный и эпидидимальный жир. «0» и «120» указывают на временные точки до и через 120 мин после нагрузки глюкозой в ГТТ.

Исследование гормональных показателей тиреоидной системы показало отсутствие значимых изменений уровня ТТГ, но уменьшение продукции fT4 и tT3, а также снижение соотношений fT4/ТТГ и tT3/ТТГ, что указывает на снижение ответа щитовидной железы на эндогенный ТТГ и признаки гипотиреоидного состояния. Обработка диабетических крыс PI4 и Comp3 не влияла на уровень ТТГ, но частично восстанавливала уровни fT4 и tT3 и их соотношение с ТТГ, в большей степени в случае PI4 (см. табл. 1). Это указывает на индуцированное PI4 и Comp3 улучшение тиреоидного статуса у крыс с СД2, что, как мы полагаем обусловлено повышением чувствительности их тканей к инсулину и лептину.

Заключение: Таким образом, разработанные нами ингибиторы фосфатаз с двойной специфичностью, имеющие структуру 4-оксо-1,4-дигидроциннолина, могут стать прототипами лекарственных препаратов для коррекции метаболических и гормональных расстройств при СД2.

Финансирование: Работа выполнена в рамках государственного задания ИЭФБ РАН № 075-00264-24-00.

Effect of tyrosine phosphatase inhibitors on hormonal parameters and insulin sensitivity in rats with diet-induced type 2 diabetes mellitus

Derkach K.*, Zorina I., Shpakov A.

Sechenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences, St. Petersburg, Russia

** derkatch_k@list.ru*

Key words: protein phosphotyrosine phosphatase 1B; T-cell protein phosphotyrosine phosphatase, type 2 diabetes; inhibitor; insulin receptor; leptin; insulin resistance

Motivation and Purpose: The key pathogenetic factors in type 2 diabetes mellitus (T2DM), the most common endocrine disease in the world leading to many complications, are insulin resistance (IR), impaired glucose tolerance and dyslipidemia. The development of pharmacological approaches aimed at normalizing insulin sensitivity, glucose and lipid homeostasis, along with diet and physical activity, is the main direction in the treatment of T2DM and its complications. One of the causes of IR is an increase in the activity of tyrosine phosphatases in response to cell stimulation with insulin, which dephosphorylate and inactivate the insulin receptor and its associated insulin receptor substrates-1 and -2 (IRS1/2), key components of insulin signaling [1, 2]. Among the phosphatases, the most important are protein phosphotyrosine phosphatase 1B (PTP1B) and T-cell protein phosphotyrosine phosphatase (TCPTP), which function as negative regulators of insulin signaling. They also suppress the leptin signaling pathways by dephosphorylating the leptin receptor and IRS1/2 and causing leptin resistance [3]. As a result, one of the ways to normalize insulin and leptin sensitivity in metabolic disorders is the use of selective phosphatase inhibitors, primarily PTP1B [4]. We have developed compounds PI4 and Comp3, inhibitors of the phosphatases PTP1B and TCPTP, created on the basis of the structure of 4-oxo-1,4-dihydrocinnoline, and it has been shown that inhibitors with dual specificity, inhibiting both PTP1B, were most effective in improving metabolic status and reducing IR, and TCPTP [5, 6]. However, their effectiveness in correcting IR and leptin resistance in T2DM has not been studied, and their effect on the parameters of the thyroid axis, the functions of which are impaired in T2DM, has not been studied. Based on this, the aim of the work was to study the effect of PI4 and Comp3 on metabolic and hormonal parameters, including the levels of thyroid-stimulating hormone (TSH) and thyroid hormones, in male rats with diet-induced T2DM.

Materials and Methods: Male Wistar rats were used in the experiments. The experiments were carried out in accordance with the requirements of the Ethics Committee of the IEPhB RAS (protocol No. 4-2/2023 of April 25, 2023). The animals were kept under standard vivarium conditions in plastic containers of 6 rats each, with free access to food and water. The synthesis of ethyl 3-(hydroxymethyl)-4-oxo-1,4-dihydrocinnoline-6-carboxylate (PI4) and 6-chloro-3-(hydroxymethyl)-cinnolin-4(1H)-one (Comp3) was carried out using reactions of Sonogashira and Richter, allowing the construction of a cinnoline framework. The target products were purified using adsorption chromatography (purity no less than 98 %), the structure of the compounds was proven using high-resolution mass spectrometry on a micrOTOF spectrometer («Bruker»,

Germany). T2DM was induced by a high-fat diet for 16 weeks, followed by treatment with streptozotocin (20 mg/kg) at week 10 to decompensate pancreatic β-cell function. Two weeks before treatment of animals with phosphatase inhibitors in the diabetic groups, animals with pronounced signs of T2DM and obesity were selected, such as an increase in body weight compared to the control by at least 10 % and impaired glucose tolerance in the glucose tolerance test (GTT), which was expressed in its level 120 min after a glucose load is not lower than 7.5 mM. Then 4 groups of rats (6 animals in each) were formed: control (C), T2DM without treatment (D), T2DM with 7-day treatment with PI4 (Dpi4) or Comp3 (Dcomp3) at a daily dose of 10 mg/kg (solution in DMSO). On day 7, blood was drawn to measure TSH and thyroid hormone levels. One day after the end of treatment, a repeated GTT was performed, assessing the levels of glucose, insulin and leptin in the blood discretely, over 120 min after the glucose load. Blood sampling was carried out under local anesthesia (lidocaine, 4 mg/kg). Glucose levels were measured using «One Touch Ultra» test strips (USA), insulin, leptin and TSH levels were measured using «ELISA for Insulin, Rat» («Cloud-Clone Corp.», USA), «ELISA for Leptin, Rat» («Cloud-Clone Corp.», USA) and «ELISA for TSH, Rat» («Cusabio Biotech Co., Ltd.», China), and the levels of free and total thyroxine (fT4, tT4) and free and total triiodothyronine (fT3, tT3) with using ELISA kits from «Immunotech» (Russia). After this, the rats were anesthetized (chloral hydrate, 400 mg/kg), decapitated, and body weight and adipose tissue were assessed. For statistical analysis, the «Microsoft Office Excel 2007» software package (USA) was used. The results were presented as $M \pm SD$. Normality of distribution was assessed using the Shapiro-Wilk test. To compare two samples with a normal distribution, Student's t-test was used; to compare a larger number of samples, analysis of variance with Bonferroni correction was used. Differences were considered significant at $p < 0.05$.

Results: Rats with T2DM had significantly increased body weight and adipose tissue, glucose levels and $AUC_{0\text{-}120}$ value, which is the integrated area under the glucose concentration curve in the GTT, as well as basal leptin levels and glucose-stimulated leptin and insulin levels (Table 1). Treatment of diabetic rats for a week with both phosphatase inhibitors resulted in a decrease in body weight and fat, and in the case of PI4 these indicators did not differ from those in controls, as well as improvement in glucose tolerance and normalization of basal and glucose-stimulated insulin and leptin levels. Insulin levels were more sensitive to treatment with phosphatase inhibitors, and the effect of PI4 on these indicators was more pronounced compared to Comp3 (Table). Thus, PTP1B and TCPTP inhibitors, characterized by dual specificity, were effective in normalizing metabolic and hormonal parameters, which is consistent with both our early results on their use for the correction of obesity [5, 6], and with the data of other authors on the restorative effect of various inhibitors of the phosphatase PTP1B on tissue sensitivity to insulin and leptin, although in this case the effects of such inhibitors attenuated with long-term treatment [7]. This may be due to the fact that suppression of PTP1B activity led to a compensatory increase in the expression and activity of TCPTP, which can to some extent replace PTP1B as a negative regulator of insulin and leptin signaling [6]. A study of hormonal parameters of the thyroid system showed the absence of significant changes in the level of TSH, but a decrease in the production of fT4 and tT3, as well as a decrease in the ratios of fT4/TSH and tT3/TSH, which indicates a decrease in the response of the thyroid gland to endogenous TSH and signs of a hypothyroid state. Treatment of diabetic rats with PI4 and Comp3 had no effect on TSH levels, but partially restored the levels of fT4 and tT3 and their ratio to TSH, to a greater

extent in the case of PI4 (Table 1). This indicates an improvement in thyroid status induced by PI4 and Comp3 in rats with T2DM, which we believe is due to an increase in the sensitivity of their tissues to insulin and leptin.

Table 1. Effect of PI4 and Comp3, endowed with dual specificity phosphatase inhibitors PTP1B and TCPTP, on body and adipose tissue weight, glucose levels and hormonal profile in rats with diet-induced T2DM

Indicator	C	D	Dpi4	Dcomp3
Body weight, g	345±9	430±11*	381±15#	393±10*#
Weight of AF, g	3.9±0.5	14.4±1.7*	7.2±0.8*#	8.1±0.7*#
Weight of EF, g	3.7±0.4	7.9±1.2	5.5±0.9	6.4±0.8*
Glucose-0, mM	4.3±0.2	6.1±0.3*	4.0±0.2#	4.6±0.8
Glucose-120, mM	5.1±0.3	9.5±1.3*	5.2±0.6#	6.2±1.4
$AUC_{0\text{-}120}$, arb.units	1109±46	1895±65*	1239±57#	1387±48*#
Insulin-0, ng/ml	0.54±0.10	0.71±0.18	0.47±0.11	0.69±0.11
Insulin-120, ng/ml	0.68±0.09	1.40±0.15*	0.77±0.10#	0.83±0.11#
Leptin-0, ng/ml	0.95±0.12	2.87±0.43*	1.77±0.18*#	1.79±0.22*#
Leptin-120, ng/ml	1.22±0.38	4.51±0.37*	2.05±0.39#	2.90±0.18*#
TSH, μIU/ml	0.56±0.09	0.71±0.12	0.50±0.05	0.58±0.07
fT4, pM	28.8±1.3	22.1±0.8*	26.9±1.1#	25.1±1.9
tT4, nM	76.2±6.9	69.1±7.0	73.8±4.6	72.2±3.3
fT3, pM	2.5±0.3	2.1±0.3	2.6±0.3	2.4±0.2
tT3, nM	3.4±0.2	2.8±0.2*	3.4±0.1#	3.2±0.3

Note. AF and EF – abdominal and epididymal fat. "0" and "120" indicate time points before and 120 min after glucose loading in the GTT.

Conclusion: Thus, the dual-specificity phosphatase inhibitors we have developed, having the structure of 4-oxo-1,4-dihydrocinnoline, can become prototypes of drugs for the correction of metabolic and hormonal disorders in T2DM.

Funding: The work is supported by the IEPhB Research Program No. 075-00264-24-00.

Список литературы/References

1. Köhn M. Turn and face the strange: a new view on phosphatases. *ACS Cent Sci*. 2020;6(4):467-477. doi 10.1021/acscentsci.9b00909
2. Liu R. et al. Human Protein Tyrosine Phosphatase 1B (PTP1B): From structure to clinical inhibitor perspectives. *Int J Mol Sci*. 2022;23(13):7027. doi 10.3390/ijms23137027
3. Zhang Z.Y. et al. Protein tyrosine phosphatases in hypothalamic insulin and leptin signaling. *Trends Pharmacol Sci*. 2015;36(10):661-674. doi 10.1016/j.tips.2015.07.003
4. Sorokoumov V.N., Shpakov A.O. Protein phosphotyrosine phosphatase 1B: Structure, function, role in the development of metabolic disorders and their correction by the enzyme inhibitors. *J Evol Biochem Physiol*. 2017;53(4):259-270. doi 10.1134/S0022093017040020
5. Деркач К.В. и др. Характеристика и биологическая активность новых ингибиторов тирозиновых фосфатаз PTP1B И TCPTP на основе 4-оксо-1,4-дигидроциннолина. *Биомедицинская химия*. 2022;68(6):427-436. doi 10.18097/PBMC20226806427
[Derkach K.V. et al. Characterization and biological activity of new 4-oxo-1,4-dihydrocinnoline-based inhibitors of the tyrosine phosphatase PTP1B and TCPTP. *Biomed Khim*. 2022;68(6):427-436. doi 10.18097/PBMC20226806427 (in Russian)]
6. Derkach K.V. et al. Dual PTP1B/TC-PTP Inhibitors: Biological Evaluation of 3-(Hydroxymethyl)cinnoline-4(*1H*)-Ones. *Int J Mol Sci*. 2023;24(5):4498. doi 10.3390/ijms24054498
7. Campos-Almazán M.I. et al. Computational methods in cooperation with experimental approaches to design protein tyrosine phosphatase 1b inhibitors in type 2 diabetes drug design: a review of the achievements of this century. *Pharmaceuticals (Basel)*. 2022;15(7):866. doi 10.3390/ph15070866

doi 10.18699/bgrs2024-9.4-05

Низкомолекулярный аллостерический агонист, хорионический гонадотропин и их комбинация как триггеры овуляции у неполовозрелых самок крыс

Деркач К., Лебедев И., Морина И., Бахтюков А., Кузнецова В., Романова И., Шпаков А.

Институт эволюционной физиологии и биохимии им. И.М. Сеченова, Российская академия наук, Санкт-Петербург, Россия

** derkatch_k@list.ru*

Ключевые слова: низкомолекулярный агонист, рецептор лютеинизирующего гормона, хорионический гонадотропин, индукция овуляции, металлопротеиназа ADAMTS-1, белок StAR

Мотивация и цель: Процесс овуляции, включающий разрыв фолликула и высвобождение яйцеклетки, реализуется путем мощного выброса гипофизом лютеинизирующего гормона (ЛГ), который воздействует на специфичные к нему рецепторы в фолликулярных клетках яичников. При проведении вспомогательных репродуктивных технологий в качестве индуктора овуляции обычно используют хорионический гонадотропин человека (ХГЧ). Однако он характеризуется рядом серьезных побочных эффектов, среди которых снижение чувствительности яичников к эндогенному ЛГ, высокие риски развития синдрома гиперстимуляции яичников [1,2]. В связи с этим ведется поиск альтернативных индукторов овуляции, и большой интерес здесь представляют низкомолекулярные агонисты рецептора ЛГ (ЛГР), которые связываются с аллостерическим сайтом, расположенным внутри трансмембранного домена, в то время как гонадотропины взаимодействуют с внеклеточным сайтом [3]. Нами разработаны агонисты ЛГР, производные тиено[2,3-d]-пиримидина, которые активируют ЛГР и его каскады. Наиболее активными среди них являются ТР3 и ТР4, которые повышают продукцию тестостерона у самцов крыс, а также усиливают синтез прогестерона и вызывают овуляцию у самок крыс [4]. Поскольку одной из возможных стратегий для индукции овуляции может служить комбинированное применение ХГЧ и низкомолекулярных агонистов ЛГР, то актуальной задачей является оценка их эффективности при совместном введении. Целью работы было сравнительное изучение эффектов ХГЧ и ТР3 на продукцию прогестерона и индукцию овуляции у неполовозрелых самок крыс при раздельном и совместном введении.

Материалы и методы: В опытах использовали самок крыс Вистар возраста 22–24 дня, эксперименты осуществляли в соответствии с требованиями Этического комитета ИЭФБ РАН (протокол № 4-1/2023 от 25 апреля 2023 г.). Для стимуляции фолликулогенеза использовали Фоллимаг (п/к, 15 МЕ/крысу) («Мосагроген», Россия). Через 48 ч после обработки Фоллимагом вводили индукторы овуляции – ТР3 (в/б, 25 мг/кг), ХГЧ (п/к, 10 МЕ/крысу) и совместно ТР3 и ХГЧ в тех же дозах. Через 8, 16 или 24 ч после введения ТР3, ХГЧ или ТР3+ХГЧ у крыс забирали влагалищные мазки для микроскопии. Через указанные временные интервалы крыс анестезировали хлоралгидратом (400 мг/кг), декапитировали, забирали

яичники, оценивали количество фолликулов и желтых тел и измеряли экспрессию целевых генов. Из шейных вен забирали кровь для определения уровня прогестерона. Исследовали 10 групп крыс (во всех n = 5): контроль (К), обработанный Фоллимагом и получавший растворители препаратов, и экспериментальные группы, которые через 48 ч после введения Фоллимага обрабатывали ТР3, ХГЧ или ТР3+ХГЧ и декапитировали через 8, 16 и 24 ч после введения ТР3 (Т8, Т16, Т24), ХГЧ (Г8, Г16, Г24) или ТР3+ХГЧ (ТГ8, ТГ16, ТГ24). Морфологический анализ яичников осуществляли, как описано ранее [4]. На срезах подсчитывали количество предовуляторных фолликулов и желтых тел, по которым оценивали число овулировавших фолликулов. Уровень прогестерона в крови определяли с помощью набора «Прогестерон-ИФА» («ХЕМА», Россия). RT-ПЦР осуществляли с помощью Applied Biosystems 7500 RT-PCR System («Life Technologies», США). Экспрессию генов *Lhr*, *Star, Adamts-1* и *Mt-1*, кодирующих ЛГР, холестерин-транспортирующий белок StAR, металлопротеиназу ADAMTS-1 и металлотионеин-1, рассчитывали методом delta-delta-C_t, используя гены актина B (*Actb*) и 18S-rRNA, как референсные. Статистический анализ проводили, используя программу «Microsoft Office Excel 2007» (США), и представляли результаты, как *M*±*SEM*. Нормальность распределения оценивали с помощью критерия Шапиро-Уилка. Значимыми считали различия при $p < 0.05$.

Результаты: Через 8 ч после введения ХГЧ и ТР3 раздельно и совместно в яичниках крыс отмечали повышение количества предовуляторных фолликулов и уровня прогестерона в крови, причем уровень прогестерона в группе ТГ8 был выше, чем в группах Г8 и Т8 (табл. 1).

Таблица 1. Количество третичных и предовуляторных фолликулов и желтых тел в яичниках и концентрация прогестерона в крови у самок крыс, обработанных Фоллимагом и через 48 ч индукторами овуляции – ХГЧ, ТР3 или ХГЧ+ТР3

Группа крыс	Третичные фолликулы, шт.	Предовуляторные фолликулы, шт.	Желтое тело, шт.	Прогестерон, нмоль/л
К	33.5 (27.5; 54.8)	2.0 (1.8; 3.2)	0.0 (0.0; 0.0)	9.1 (5.4; 15.3)
Г8	31.5 (27.0; 40.8)	14.0 (8.5; 20.2)[a]	0.0 (0.0; 0.0)	47.7 (32.3; 66.4)[ab]
Г16	25.0 (19.8; 34.5)	1.5 (0.8; 3.0)	2.5 (2.0; 3.3)[a]	18.4 (14.2; 27.2)
Г24	37.5 (24.8; 44.2)	0.0 (0.0; 1.2)	5.5 (3.0; 6.3)[a]	6.7 (6.0; 9.9)[b]
Т8	30.5 (25.8; 45.0)	11.5 (10.5; 13.5)[a]	0.0 (0.0; 0.0)	25.2 (15.4; 29.6)[b]
Т16	23.0 (21.2; 25.8)	1.5 (0.8; 3.0)	3.0 (2.0; 3.5)[a]	18.7 (13.0; 27.1)
Т24	35.0 (27.2; 44.0)	0.0 (0.0; 1.0)[ab]	4.5 (2.8; 6.3)[a]	6.2 (4.0; 9.5)[b]
ТГ8	18.0 (14.8; 26.8)	8.5 (5.8; 12.2)[a]	0.0 (0.0; 0.0)	84.3 (63.5; 88.6)[a]
ТГ16	15.5 (12.5; 20.0)[a]	2.0 (1.0; 2.2)	5.0 (3.8; 7.0)[a]	25.8 (18.9; 28.6)
ТГ24	25.5 (19.2; 47.5)	3.0 (1.0; 4.5)	5.5 (3.5; 6.3)[a]	19.2 (17.2; 28.9)

Примечание. Различия значимы: [a] – по сравнению с К, [b] – при сравнении групп Г8 или Т8 с группой ТГ8, Г16 или Т16 с ТГ16, Г24 или Т24 с ТГ24 при $p<0.05$.

Через 16 и 24 ч в группах, обработанных ХГЧ и ТР3 раздельно и совместно, число предовуляторных фолликулов снижалось, что сопровождалось снижением уровня прогестерона (табл. 1). Оценка экспрессии овариальных генов показала, что через 8 ч после обработки ХГЧ или ТР3 уровень мРНК ЛГР повышался, в то время как в группе ТГ8 этого не происходило (табл. 2). Через 16 ч в группах Г16 и ТГ16

экспрессия гена *Lhr* снижалась ниже ее уровня в контроле, в группе Т16 она от контроля не отличалась. Через 24 ч экспрессия гена *Lhr* восстанавливалась, наиболее отчетливо в группе ТГ24. Экспрессия гена *Star*, кодирующего ключевой стероидогенный белок StAR, и гена металлопротеиназы ADAMTS-1, важнейшего маркера начальных стадий индукции овуляции, была максимальной через 8 ч, причем в группе ТГ8 она была выше, чем в группах Г8 и Т8, а затем с различной интенсивностью снижалась (табл. 2). Обратная картина отмечалась для экспрессии гена металлотионеина-1, маркера формирования желтых тел, которая была максимальна через 24 ч после обработки.

Таблица 2. Экспрессия генов, кодирующих ЛГР (*Lhr*), белок StAR (*Star*), металлопротеиназу ADAMTS-1 (*Adamts-1*) и металлотионеин-1 (*Mt-1*) в яичниках самок крыс, обработанных Фоллимагом и индукторами овуляции

Группа крыс	*Lhr*, отн ед.	*Star*, отн ед.	*Adamts-1*, отн ед.	*Mt-1*, отн ед.
К	0.96 (0.86; 1.21)	1.17 (0.84; 1.22)	1.01 (0.76; 1.28)	0.98 (0.82; 1.27)
Г8	3.62 (2.52; 4.54)[ab]	13.63 (10.82; 15.41)[ab]	7.83 (5.42; 10.58)[ab]	1.48 (0.99; 2.41)
Г16	0.28 (0.15; 0.40)[a]	12.05 (10.86; 14.86)[a]	2.81 (1.97; 3.98)[a]	5.10 (4.12; 6.70)[a]
Г24	0.45 (0.36; 0.68)	9.75 (7.82; 11.39)[a]	2.55 (2.38; 3.52)[a]	7.99 (6.54; 8.55)[a]
Т8	3.15 (2.27;3.80)[ab]	11.46 (9.18; 14.68)[ab]	6.73 (5.02; 7.66)[ab]	1.53 (1.16; 2.55)
Т16	0.84 (0.42; 1.45)	5.86 (2.39; 9.00)[a]	2.42 (2.30; 2.89)[a]	3.71 (2.52; 4.55) [a]
Т24	1.29 (0.85; 2.68)	3.19 (2.38; 4.42)[ab]	3.14 (2.00; 3.37)	4.97 (3.40; 5.99)[a]
ТГ8	1.40 (1.17; 1.48)	20.23 (19.26; 22.42)[a]	16.34 (12.63; 29.97)[a]	1.84 (1.28; 2.95)
ТГ16	0.40 (0.14; 0.61)[a]	11.57 (10.16; 13.62)[a]	1.98 (1.78; 3.06)[a]	5.86 (3.81; 7.02) [a]
ТГ24	1.31 (0.86; 1.56)	7.18 (5.59; 8.68)[a]	3.68 (2.81; 4.13)[a]	7.54 (5.08; 9.41)[a]

Примечание. Обозначения значимости различий те же, что и в табл. 1.

Выводы: ХГЧ и разработанный нами низкомолекулярный аллостерический ЛГР-агонист ТР3 как по отдельности, так и совместно вызывали овуляцию у неполовозрелых самок крыс, стимулировали у них продукцию прогестерона, повышали экспрессию генов, вовлеченных в овариальный стероидогенез и индукцию овуляции. При этом комбинация препаратов была более эффективной в отношении стимуляции синтеза прогестерона и экспрессии генов стероидогенного белка StAR и металлопротеиназы ADAMTS-1, маркера овуляции, что было ассоциировано с более быстрым образованием желтых тел. Тем самым, при совместном введении ХГЧ и ТР3 их активность, как индукторов овуляции, повышается, и этот феномен нуждается в дальнейшем изучении.

Финансирование: Исследование поддержано грантом Российского Научного Фонда (проект № 19-75-20122).

Low-molecular-weight allosteric agonist, human chorionic gonadotropin and their combination as triggers of ovulation in immature female rats

Derkach K., Lebedev I., Morina I., Bakhtyukov A., Kuznetsova V., Romanova I., Shpakov A.

Sechenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences, St. Petersburg, Russia

** derkatch_k@list.ru*

Key words: low molecular weight agonist, luteinizing hormone receptor, human chorionic gonadotropin, ovulation induction, ADAMTS-1 metalloproteinase, StAR protein

Motivation and Purpose: The process of ovulation, which includes rupture of the follicle and release of the egg, is realized by the significant release of luteinizing hormone (LH) by the pituitary, which acts on its specific receptors in the follicular cells of the ovaries. When carrying out assisted reproductive technologies, human chorionic gonadotropin (hCG) is usually used as ovulation inducers. However, it is characterized by a number of serious side effects, including a decrease in the sensitivity of the ovaries to endogenous LH and a high risk of ovarian hyperstimulation syndrome [1,2]. In this regard, the search for alternative ovulation inducers is underway, and much interest is focused on low-molecular-weight LH receptor (LHR) agonists, which bind to an allosteric site located within the transmembrane domain, while gonadotropins interact with the extracellular site [3]. We have developed LHR agonists, thieno[2,3-d]-pyrimidine derivatives, which activate LHR and its cascades. The most active among them are TP3 and TP4, which increase testosterone production in male rats, and enhance progesterone synthesis and induce ovulation in female rats [4]. Since one of the possible strategies for inducing ovulation may be the combined use of hCG and low-molecular-weight LHR agonists, an urgent task is to assess their effectiveness when administered together. The purpose of the work was a comparative study of the effects of hCG and TP3 on progesterone production and ovulation induction in immature female rats when administered separately and together.

Materials and methods: Female Wistar rats aged 22–24 days were used in the experiments. The experiments were carried out in accordance with the requirements of the Ethics Committee of the IEPhB RAS (protocol No. 4-1/2023 of April 25, 2023). To stimulate folliculogenesis, Follimag (s.c., 15 IU/rat) ("Mosagrogen", Russia) was used. 48 hours after treatment with Follimag, ovulation inducers were administered: TP3 (i.p., 25 mg/kg), hCG (s.c., 10 IU/rat) and together TP3 and hCG in the same doses. At 8, 16, or 24 h after administration of TP3, hCG, or TP3+hCG, vaginal smears were collected from rats for microscopy. At the indicated time intervals, rats were anesthetized with chloral hydrate (400 mg/kg), decapitated, ovaries were harvested, follicles and corpus luteum numbers were assessed, and the expression of target genes was measured. Blood samples were collected from the neck veins to determine progesterone levels. We studied 10 groups of rats (all $n = 5$): control (C), treated with Follimag and receiving drug solvents, and the experimental groups, which 48 h after administration of Follimag were treated with TP3, hCG or TP3+hCG and decapitated after 8, 16 and 24 h after administration of TP3 (T8, T16, T24), hCG (G8, G16, G24) or TP3+hCG (TG8, TG16,

TG24). Morphological analysis of the ovaries was carried out as described previously [4]. On the sections, the number of preovulatory follicles and corpus luteum was counted, from which the number of ovulated follicles was estimated. The level of progesterone in the blood was determined using the Progesterone-ELISA kit ("CHEMA", Russia). RT-PCR was performed using an Applied Biosystems 7500 RT-PCR System ("Life Technologies", USA). The expression of the *Lhr*, *Star*, *Adamts-1* and *Mt-1* genes encoding LHR, cholesterol transporting protein StAR, metalloproteinase ADAMTS-1 and metallothionein-1 was calculated by the delta-delta-Ct method, using actin B (*Actb*) and 18S-rRNA genes as reference genes. Statistical analysis was carried out using "Microsoft Office Excel 2007" (USA), and the results were presented as *M*±*SEM*. Normality of distribution was assessed using the Shapiro-Wilk test. Differences were considered as significant at $p < 0.05$.

Results: 8 hours after the administration of hCG and TP3 separately and together, an increase in the number of preovulatory follicles and the level of progesterone in the blood was shown in the ovaries of rats, and the level of progesterone in the TG8 group was higher than in the G8 and T8 groups (Table 1). After 16 and 24 hours, in the groups treated with hCG and TP3 separately and together, the number of preovulatory follicles decreased, which was accompanied by a decrease in progesterone levels (Table 1).

Table 1. The number of tertiary and preovulatory follicles and corpus luteum in the ovaries and the concentration of progesterone in the blood in female rats treated with Follimag and after 48 hours with ovulation inducers - hCG, TP3 or hCG+TP3

Rat group	Tertiary follicles, units	Preovulatory follicles, units	Corpus luteum, units	Progesterone, nmol/l
C	33.5 (27.5; 54.8)	2.0 (1.8; 3.2)	0.0 (0.0; 0.0)	9.1 (5.4; 15.3)
G8	31.5 (27.0; 40.8)	14.0 (8.5; 20.2)[a]	0.0 (0.0; 0.0)	47.7 (32.3; 66.4)[ab]
G16	25.0 (19.8; 34.5)	1.5 (0.8; 3.0)	2.5 (2.0; 3.3)[a]	18.4 (14.2; 27.2)
G24	37.5 (24.8; 44.2)	0.0 (0.0; 1.2)	5.5 (3.0; 6.3)[a]	6.7 (6.0; 9.9)[b]
T8	30.5 (25.8; 45.0)	11.5 (10.5; 13.5)[a]	0.0 (0.0; 0.0)	25.2 (15.4; 29.6)[b]
T16	23.0 (21.2; 25.8)	1.5 (0.8; 3.0)	3.0 (2.0; 3.5)[a]	18.7 (13.0; 27.1)
T24	35.0 (27.2; 44.0)	0.0 (0.0; 1.0)[ab]	4.5 (2.8; 6.3)[a]	6.2 (4.0; 9.5)[b]
TG8	18.0 (14.8; 26.8)	8.5 (5.8; 12.2)[a]	0.0 (0.0; 0.0)	84.3 (63.5; 88.6)[a]
TG16	15.5 (12.5; 20.0)[a]	2.0 (1.0; 2.2)	5.0 (3.8; 7.0)[a]	25.8 (18.9; 28.6)
TG24	25.5 (19.2; 47.5)	3.0 (1.0; 4.5)	5.5 (3.5; 6.3)[a]	19.2 (17.2; 28.9)

Note. The differences are significant: [a] – compared with C, [b] – when comparing groups G8 or T8 with group TG8, G16 or T16 with TG16, and G24 or T24 with TG24 at $p<0.05$.

Assessment of ovarian gene expression showed that 8 hours after treatment with hCG or TP3, the level of LHR mRNA increased, while this did not happen in the TG8 group (Table 2). After 16 hours, in the T16 and TG16 groups, the expression of the *Lhr* gene decreased below its level in the control, and in the T16 group it did not differ from the control. After 24 hours, *Lhr* gene expression was restored, most pronounced in the TG24 group. Expression of the *Star* gene, encoding the key steroidogenic protein StAR, and the metalloproteinase gene ADAMTS-1, the most important marker of the initial stages of ovulation induction, was maximum after 8 hours, and in the TG8 group it was higher than in the G8 and T8 groups, and then with varying intensity decreased (Table 2). The opposite pattern was observed for the expression of the metallothionein-1 gene, a marker of corpus luteum formation, which was maximum 24 hours after treatment.

Table 2. Expression of genes encoding LHR (*Lhr*), StAR protein (*Star*), metalloproteinase ADAMTS-1 (*Adamts-1*) and metallothionein-1 (*Mt-1*) in the ovaries of female rats treated with Follimag and ovulation inducers

Rat group	*Lhr*, arb.units	*Star*, arb.units	*Adamts-1*, arb.units	*Mt-1*, arb.units
C	0.96 (0.86; 1.21)	1.17 (0.84;1.22)	1.01 (0.76; 1.28)	0.98 (0.82; 1.27)
G8	3.62 (2.52; 4.54)[ab]	13.63 (10.82; 15.41)[ab]	7.83 (5.42; 10.58)[ab]	1.48 (0.99; 2.41)
G16	0.28 (0.15; 0.40)[a]	12.05 (10.86; 14.86)[a]	2.81 (1.97; 3.98)[a]	5.10 (4.12; 6.70)[a]
G24	0.45 (0.36; 0.68)	9.75 (7.82; 11.39)[a]	2.55 (2.38; 3.52)[a]	7.99 (6.54; 8.55)[a]
T8	3.15 (2.27; 3.80)[ab]	11.46 (9.18; 14.68)[ab]	6.73 (5.02; 7.66)[ab]	1.53 (1.16; 2.55)
T16	0.84 (0.42; 1.45)	5.86 (2.39; 9.00)[a]	2.42 (2.30; 2.89)[a]	3.71 (2.52; 4.55) [a]
T24	1.29 (0.85; 2.68)	3.19 (2.38; 4.42)[ab]	3.14 (2.00; 3.37)	4.97 (3.40; 5.99)[a]
TG8	1.40 (1.17; 1.48)	20.23 (19.26; 22.42)[a]	16.34 (12.63; 29.97)[a]	1.84 (1.28; 2.95)
TG16	0.40 (0.14; 0.61)[a]	11.57 (10.16; 13.62)[a]	1.98 (1.78; 3.06)[a]	5.86 (3.81; 7.02) [a]
TG24	1.31 (0.86; 1.56)	7.18 (5.59; 8.68)[a]	3.68 (2.81; 4.13)[a]	7.54 (5.08; 9.41)[a]

Note. The designations for the significance of differences are the same as in Table 1.

Conclusion: hCG and the low-molecular-weight allosteric LHR agonist TP3, which we developed, both individually and together, induced ovulation in immature female rats, stimulated progesterone production, and increased the expression of genes involved in ovarian steroidogenesis and ovulation induction. Moreover, the combination of drugs was more effective in stimulating the synthesis of progesterone and the expression of the genes for the steroidogenic protein StAR and metalloproteinase ADAMTS-1, an ovulation marker, which was associated with more rapid formation of the corpus luteum. Thus, with the joint administration of hCG and TP3, their activity as ovulation inducers increases, and this phenomenon requires further study.

Funding: The study is supported by a grant from the Russian Science Foundation (project No. 19-75-20122).

Список литературы/References

1. Riccetti L. et al. Human Luteinizing Hormone and Chorionic Gonadotropin Display Biased Agonism at the LH and LH/CG Receptors. *Sci. Rep*. 2017;7(1):940. doi: 10.1038/s41598-017-01078-8
2. Singh R. et al. Gonadotropins as pharmacological agents in assisted reproductive technology and polycystic ovary syndrome. *Trends Endocrinol. Metab*. 2023;34(4):194-215. doi: 10.1016/j.tem.2023.02.002
3. Lazzaretti C. et al. Allosteric modulation of gonadotropin receptors. *Front. Endocrinol. (Lausanne).* 2023;14:1179079. doi: 10.3389/fendo.2023.1179079
4. Derkach K.V. et al. Comparison of Steroidogenic and Ovulation-Inducing Effects of Orthosteric and Allosteric Agonists of Luteinizing Hormone/Chorionic Gonadotropin Receptor in Immature Female Rats. *Int. J. Mol. Sci*. 2023;24(23):16618. doi: 10.3390/ijms242316618

doi 10.18699/bgrs2024-9.4-06

Скрининг молекулярных свойств и аффинности связывания компонентов лавра благородного (*Laurus nobilis*) с ферментами пищеварения человека

Иванов Н.*, Пенцак Е.

Институт органической химии им. Н.Д. Зелинского РАН, Москва, Россия

** nikstar.ivanov@yandex.ru*

Ключевые слова: лавр благородный; липаза; биологически активные вещества

Мотивация и цель: Биологически активные вещества лавра благородного (*Laurus nobilis*) отличаются высокими антидиабетическими свойствами, имеют высокий потенциал применения в эндокринологической практике [1–5], однако информация о молекулярном взаимодействии некоторых компонентов лавра благородного с молекулами ферментами человека представляется недостаточной. В связи с этим целью данной работы является изучение молекулярных свойств и аффинности связывания компонентов лавра благородного с трехмерными структурами пищеварительных ферментов человека.

Методы и алгоритмы: Трехмерная структура лизосомальной кислой липазы человека была получена методом гомологичного моделирования в предыдущей работе [6]. Процедура молекулярного докинга данного фермента с трехмерными структурами биологически активных компонентов лавра благородного [7] была проведена с использованием программного обеспечения AutoDock Vina, Proteins.Plus.

Результаты: в результате проведенных исследований получены данные аффинности связывания и типа молекулярного взаимодействия основных компонентов лавра благородного с лизосомальной кислой липазой человека, представленные в табл. 1.

Таблица 1. Данные аффинности связывания и типа молекулярного взаимодействия

Компоненты	Аффинность связывания, Ккал/моль	Тип взаимодействия
1,8-цинеол	–8.8	Водородное
терпинилацетат	–7.9	Гидрофобное
α-пинен	–6.8	Гидрофобное

Выводы: Экспериментальные данные *in silico* исследования показывают наличие белок-лигандного взаимодействия между лизосомальной кислой липазой человека и компонентами экстракта лавра благородного. Одним из направлений дальнейших исследований может быть скрининг *in vitro* кинетики связывания компонентов растения и ферментов пищеварения человека.

Screening of molecular properties and binding affinities of laurel (*Laurus nobilis*) components with human digestive enzymes

Ivanov N.*, Pentsak E.

N.D. Zelinsky Institute of Organic Chemistry of the Russian Academy of Sciences, Moscow, Russia

** nikstar.ivanov@yandex.ru*

Key words: bay laurel; lipase; biologically active substances

Motivation and Aim: The biologically active substances of laurel nobilis (*Laurus nobilis*) are distinguished by high antidiabetic properties and have a high potential for use in endocrinological practice [1–5], however, information on the molecular interaction of some components of laurel nobilis with human enzyme molecules seems insufficient. In this regard, the purpose of this work is to study the molecular properties and binding affinities of *Laurus nobilis* components with the three-dimensional structures of human digestive enzymes.

Methods and Algorithms: The three-dimensional structure of human lysosomal acid lipase was obtained by homology modeling in a previous work [6]. The procedure for molecular docking of this enzyme with three-dimensional structures of biologically active components of *Laurus nobilis* [7] was carried out using AutoDock Vina, Proteins.Plus software.

Results: As a result of the studies, data on the binding affinity and type of molecular interaction of the main components of *Laurus nobilis* with human lysosomal acid lipase were obtained, presented in the Table 1.

Table 1. Affinity and type of molecular interaction

Components	Binding affinity, Kcal/mol	Interaction
1,8-cineol	–8.8	Hydrogen
terpinyl acetate	–7.9	Hydrophobic
α-pinene	–6.8	Hydrophobic

Conclusion: Experimental data from *in silico* studies indicate the presence of protein-ligand interactions between human lysosomal acid lipase and components of bay laurel extract. However, further *in vitro* studies are needed to determine how these components can be used.

Список литературы/References

1. Mohammed R.R. et al. Biomedical effects of Laurus nobilis L. leaf extract on vital organs in streptozotocin-induced diabetic rats: Experimental research. *Ann Med Surg*. 2021;61:188-197
2. Daher C. et al. Laurus nobilis leaves extract protects against high fat diet-induced type 2 Diabetes in rats. *J Pharmacogn Phytotherapy*. 2021;13(3):82-90
3. Bourebaba N. et al. *Laurus nobilis* ethanolic extract attenuates hyperglycemia and hyperinsulinemia-induced insulin resistance in HepG2 cell line through the reduction of oxidative stress and improvement of mitochondrial biogenesis – possible implication in pharmacotherapy. *Mitochondrion*. 2021;59:190-213

4. Senou M., Lokonon J.E., Ayitchehou G. et al. Antidiabetic activity of aqueous extracts of laurus nobilis, a spice used by beninese traditional therapists. *Am J Med Sci*. 2021;9:115-119
5. Chbili C., Maoua M., Selmi M. et al. Evaluation of daily Laurus nobilis tea consumption on lipid profile biomarkers in healthy volunteers. *J Am College of Nutrition*. 2020;39(8):733-738
6. Иванов Н.В. Молекулярный *in silico* скрининг и докинг потенциальных ингибиторов активности ферментов растительного сырья. *Хранение и переработка сельхозсырья*. 2023;(1):117-135. doi 10.36107/spfp.2023.399
[Ivanov N.V. Molecular in silico screening and docking of potential inhibitors of enzyme activity of plant raw materials. *Storage and Processing of Farm Products*. 2023;(1):117-135. doi 10.36107/spfp.2023.399 (in Russian)]
7. Fidan H. et al. Chemical composition and antimicrobial activity of Laurus nobilis L. essential oils from Bulgaria. *Molecules*. 2019;24(4):804

doi 10.18699/bgrs2024-9.4-07

Влияние ионов железа на термическую устойчивость молекулы ТПГ2 с мутацией P447R

Комлева П.Д.[1]*, Терентьева Е.И.[2], Дееб Р.[2], Куликов А.В.[1]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

** polina.komleva@gmail.com*

Ключевые слова: триптофангидроксилаза 2; серотониновая система; полиморфизм C1473G; фармакологические шапероны; ионы железа; температурная стабильность

Мотивация и цель: Фермент триптофангидроксилаза 2 (ТПГ2) катализирует первую и ключевую стадию синтеза серотонина (5-НТ) в присутствии кофактора, тетрагидробиоптерина, ионов Fe^{2+} и кислорода [1]. Полиморфизм C1473G приводит к замене P447R в каталитическом домене фермента, что снижает активность фермента вдвое [2, 3]. Ранее было показано, что мутация нарушает 3D укладку молекулы и снижает стабильность фермента, что отражается на времени жизни молекулы [4, 5]. Фармакологические шапероны – это небольшие молекулы, способные специфически связываться с мутантными белковыми молекулами и восстанавливать их 3D-структуру. Ранее было доказано, что BH_4 способен увеличивать температуру инактивации (T_{50}) фермента *in vitro*. Однако BH_4 не позволяет восстановить температурную стабильность ТПГ2 до уровня дикого типа. Известно, что ионы двухвалентного железа участвуют в механизме ферментативного гидроксилирования L-триптофана и, следовательно, выполняют каталитическую функцию [1]. Способности ионов двухвалентного железа стабилизировать молекулу ТПГ2 не были изучены. Таким образом, цель данной работы – оценить эффект ионов Fe^{2+} и Fe^{3+} на термическую устойчивость ТПГ2 у мышей с мутацией P447R.

Методы и алгоритмы: Исследование проводилось на половозрелых самцах мышей линий C57BL/6 (n = 15) и Balb/c (n = 15), несущих1473C и 1473G аллели соотвественно. Из среднего мозга выделяли ТПГ2, затем чистый экстракт от мышей одного генотипа объединяли и разделяли на порции 300 мкл. Концентрации рабочих растворов $FeSO_4$ и $FeCl_3$ составляли 0.01, 0.05, 0.2 мМ. Для определения температурной стабильности 7.5 мкл экстракта ТПГ2 смешивали с 7.5 мкл буфера или одного из рабочих растворов, затем смесь прогревали в течение 2 мин при 48, 50, 52, 54, 56, 58, 60, 62, и 64 °C. Активность фермента определяли по скорости синтеза 5-гидрокситриптофана (5-HTP) после инкубации в течение 15 мин при 37 °C в присутствии L-триптофана (0.4 мМ) и кофактора 6-метил-5,6,7,8-тетрагидроптеридин (0.3 мМ). Уровень 5-HTP определяли с помощью ВЭЖХ с электрохимическим детектором. Температурную стабильность оценивали по температуре, инкубация при которой снижает активность фермента вдвое (T_{50}).

Результаты: Показано влияние факторов «генотип» ($F(1,28) = 648.7$, $p < 0.001$) и «препарат» ($F(3,28) = 192.5$, $p < 0.001$), но не их взаимодействия ($F(3,28) < 1$) на изменчивость T_{50} при воздействии ионами двухвалентного железа. Все три концентрации $FeSO_4$ повышают T_{50} для молекул ТПГ2 как мутантного, так и

дикого типа. При этом концентрации 0.05 и 0.2 мМ $FeSO_4$ повышают величину T_{50} для мутантных молекул ТПГ2 до уровня T_{50} для молекул ТПГ2 дикого типа.
Для ионов Fe^{3+} было показано высокое влияние факторов «генотип» ($F(1,34) = 714.2$, $p < 0.001$) и «препарат» ($F(3,34) = 86.5$, $p < 0.001$), но не их взаимодействия ($F(3,34) = 1.6$, $p = 0.21$) на изменчивость T_{50}. Все три концентрации $FeCl_3$ повышают T_{50} для молекул ТПГ2 как мутантного, так и дикого типа. При этом ни одна из исследованных концентрации $FeCl_3$ не повышает величину T_{50} для мутантных молекул ТПГ2 до уровня T_{50} для молекул ТПГ2 дикого типа.
Увеличение T_{50} (ΔT_{50}), вызванное двухвалентным железом (Balb/c, +4.12±0.2 °C; C57BL/6,+3.82±0.21 °C) было выше для мутантной ТПГ2 по сравнению с трехвалетным железом (Balb/c, +3.12±0.21 °C; C57BL/6,+3.63±0.21 °C).
Выводы: Соли двух- и трехвалентного железа одинаково хорошо увеличивают T_{50} ТПГ2 дикого типа. В то же время, ионы двух валентного железа сильнее увеличивают T_{50} мутантной ТПГ2 по сравнению с ионами трехвалентного железа. Таким образом, впервые было показано, что ионы железа, помимо каталитической функции, способны стабилизировать молекулы ТПГ2.
Финансирование: Исследование поддержано грантом РНФ № 24-15-00078.

Ferrous ions affect thermal stability of a TPH2 molecule carrying a P447R mutation

Komleva P.D.[1]*, Terentieva E.I.[2], Deeb R.[2], Kulikov A.V.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *polina.komleva@gmail.com*

Key words: tryptophan hydroxylase 2; serotonin system; C1473G polymorphism; pharmacological chaperons; ferrous ions; thermal stability

Motivation and Aim: Tryptophan hydroxylase 2 (TPH2) is an enzyme that catalyzes the first key stage of serotonin (5-HT) synthesis in presence of a cofactor, tetrahydrobiopterin (BH_4), Fe^{2+} ions and oxygen [1]. C1473G polymorphism leads to a P447R substitution in a catalytic domain of the enzyme, that decreases enzyme activity twofold [2, 3]. Previously it had been shown that the mutation disturbs molecular 3D structure and decreases stability of the enzyme, which reflects molecular lifetime [4, 5]. Pharmacological chaperons are small molecules that can specifically bind to mutant molecules and restore their folding. Recently it was indicated that BH_4 can be used to increase the temperature of the enzyme's thermal inactivation (T_{50}) in vitro. However, BH_4 does not increase TPH2's thermal stability to the level of a wild type enzyme. Divalent ferrous ions are involved in the mechanism of enzymatic hydroxylation of L-tryptophan and, therefore, perform a catalytic function [1]. Stabilizing abilities of ferrous ions have not yet been studied when applied to TPH2. Therefore, the aim of this study is to evaluate an effect of Fe^{2+} and Fe^{3+} ions on thermal stability of TPH2 in mice carrying a P447R mutation.
Methods and Algorithms: The study was conducted on mature male C57Bl ($n = 15$) and Balb/c ($n = 15$) mice, carrying 1473C and 1473G alleles respectively. TPH2 was extracted from their midbrain, then purified supernatant from the mice of the same

genotype was mixed together and aliquoted in 300 mkl portions. $FeSO_4$ and $FeCl_3$ solutions were diluted to the final concentrations of 0.01, 0.05 and 0.2 mM. To determine thermal stability, 7.5 mkl of TPH2 extract were mixed with 7.5 mkl of a buffer or an experimental solution, then the mix was heated for 2 min at 48, 50, 52, 54, 60, 62 and 64 °C. Enzyme activity was evaluated by the speed of 5-hydroxytryptophan (5-HTP) synthesis after 15 min of incubation at 37 °C in the presence of L-tryptophan (0.4 mM) and a cofactor, 6-methyl-5,6,7,8-tetrahydroptheridin (0.3 mM). 5-HTP level was determined using HLPC with an electrochemical detector. Thermal stability was evaluated by the temperature, at which enzyme activity decreases twofold (T_{50}).

Results: Marked effect of "genotype" ($F(1,28) = 648.7$, $p < 0.001$) and "treatment" ($F(3,28) = 192.5$, $p < 0.001$), but not their interaction ($F(3,28) < 1$) on T_{50} value, when treated with divalent ferrous ions, were shown. All three concentrations of $FeSO_4$ increased T_{50} for both mutant and wild type TPH2.

As for Fe^{3+} ions, significant influence of "genotype" ($F(1,34) = 714.2$, $p < 0.001$) and "treatment" ($F(3,34) = 86.5$, $p < 0.001$) but no their interaction ($F(3,28) < 1$) on T_{50} were shown. All three concentrations of $FeCl_3$ increased T_{50} for both mutant and wild type TPH2. At the same time, none of the studied $FeCl_3$ concentrations raises T_{50} value to the level of wild type TPH2.

T_{50} increase (ΔT_{50}), induced by divalent iron (Balb/c, +4.12±0.2 °C; C57BL/6,+3.82±0.21°C) was greater in mutant TPH2, compared to trivalent iron effect (Balb/c, +3.12±0.21 °C; C57BL/6,+3.63±0.21 °C).

Conclusion: Divalent and trivalent iron ions increase T_{50} value similarly well for wild type TPH2. However, divalent ferrous ions have a significantly more pronounced effect on a mutant TPH2 in comparison to trivalent ions.

Thereby, it was first shown that ferrous ions are capable of stabilizing TPH2 molecules, besides performing their catalysis function.

Funding: The study is supported by RSF No. 24-15-00078.

Список литературы/References

1. Fitzpatrick P.F. The aromatic amino acid hydroxylases: Structures, catalysis, and regulation of phenylalanine hydroxylase, tyrosine hydroxylase, and tryptophan hydroxylase. *Arch Biochem Biophys*. 2023;735:109518. doi 10.1016/j.abb.2023.109518
2. Zhang X., Beaulieu J.M., Sotnikova T.D., Gainetdinov R.R., Caron M.G. Tryptophan hydroxylase-2 controls brain serotonin synthesis. *Science*. 2004;305:217. doi 10.1126/science.1097540
3. Kulikov A.V., Osipova D.V., Naumenko V.S., Popova N.K. Association between Tph2 gene polymorphism, brain tryptophan hydroxylase activity and aggressiveness in mouse strains. *Genes Brain Behav*. 2005;4:482-485. doi 10.1111/j.1601-183X.2005.00145.x
4. Arefieva A.B., Komleva P.D., Naumenko V.S., Khotskin N.V., Kulikov A.V. *In vitro* and *in vivo* chaperone effect of (R)-2-amino-6-(1R,2S)-1,2-dihydroxypropyl)-5,6,7,8-tetrahydropterin-4(3H)-one on the C1473G mutant tryptophan hydroxylase 2. *Biomolecules*. 2023;13:1458. doi 10.3390/biom13101458
5. Arefieva A.B., Kulikov A.V. Comparison of fluorometric and chromatographic methods of in vitro assay of tryptophan hydroxylase 2, the key enzyme of serotonin synthesis in the brain. *Bull Exp Biol Med*. 2023;174:509-513. doi 10.1007/s10517-023-05738-w

doi 10.18699/bgrs2024-9.4-08

Изучение первичных фармакодинамических эффектов и возможного цитотоксического действия синтетического пептидомиметика КАМП-1

Любушкина Е.М.*, Бондарева Е.А., Солдатова М.С., Мадонов П.Г.
Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия
** lizaveta.lyubushkina@list.ru*

Ключевые слова: антибиотикорезистентность; КАМП-1; МИК; острая токсичность

Мотивация и цель: Антибактериальная резистентность представляет собой угрозу для общественного здравоохранения, по подсчетам команды под руководством O'Neil к 2050 году смертность достигнет до 10 млн человек в год, а финансовые потери приведут к убыткам более 100 трлн долларов [1]. Современные технологии антибактериальной терапии реализуется за счет несколько групп антибиотиков, к большинству из которых уже сформировалась резистентность. Поиск новых антибактериальных средств, к которым еще не сформировалась резистентность, является эффективным решением данной проблемы. На сегодняшний день синтетические антимикробные пептидомиметики представляют интерес из-за отсутствия специфической микробной мишени, а, следовательно, ниже риск развития резистентности к ним. Целью работы было определить параметры антибактериальной и противогрибковой активностей, а также детальное изучение острой токсичности инновационного пептидомиметика (КАМП-1).
Методы и алгоритмы: КАМП-1 состоит из остатка аргинина, триптофана модифицированного трет-бутильными группами, и еще одного аргинина с фенэтиламидной группой. Молекулярная масса пептида составляет 788 Дальтон. Для тестирования противомикробной активности использовались тест-культуры аттестационных штаммов бактерий и грибов. Применялся метод серийных разведений КАМП-1 в жидкой питательной среде. В кратный ряд разведений вещества добавляли взвесь инокулюма тестируемой культуры. После инкубации делали высев на плотную питательную среду с последующим подсчетом КОЕ. Наименьшая доза вещества, при которой наблюдалось полное подавление роста культуры бактерий, считалась минимальной ингибирующей концентрацией (МИК). В каждой исследуемой группе по результатам повторных опытов рассчитывали средний показатель и стандартную ошибку. Полученные результаты обрабатывались с использованием пакета статистических программ Biostatistics. В исследованиях *in vitro* МИК изучаемого вещества для тест-штаммов не превышал 100 мг/л, что согласно методическим рекомендациям по доклиническим исследованиям лекарственных веществ, характеризует его перспективным противомикробным средством [2]. В дальнейшем определялись параметры острой токсичности после однократного введения в ретроорбитальный синус и ежедневного наблюдения за животными в течение 14 дней. Использовались беспородные половозрелые мыши обоих полов. Первая группа мышей была получена и выращена в конвенциональных условиях

(n = 40), а вторая группа мышей – в условиях SPF-вивария (n = 48). Главной отличительной особенностью являлось наличие микробиома у конвенциональных мышей с целью изучения различий в результатах между нестерильными и стерильными испытуемыми, содержащиеся в условиях SPF-вивария. В первую и вторую группы были включены мыши линии CD-1 обоих полов с внешними хорошими показателями экстерьера и весовом диапазоне от 25–35 г. Животные содержались в стандартных условиях ухода со свободным доступом к пище и воде. Для изучения острой токсичности исследуемой субстанции методом пробит-анализа по В.Б. Прозоровскому высчитывались летальные дозы LD_{10}, LD_{50}, LD_{90} при помощи программы StatPlus. Раствор КАМП-1 был изготовлен в разных концентрациях от 7,0–10 мг/кг и вводили животным из расчета 1 мкл на 1 г веса. Павшим животным проводили некроскопическое исследование, а по прошествии 14 дней выжившие животные подвергались эвтаназии при помощи закиси углерода и далее осуществлялось макроскопическое исследование внутренних органов.

Результаты: Полученные данные демонстрируют, что КАМП-1 обладает антимикробной и противогрибковой активностью, что отображено в табл. 1. МИК КАМП-1 меньше допустимого предела в 100 мг/л (100 мкг/мл) на изученных аттестованных тест-культурах: *S. aureus*, *B. cereus*, *Corynebacterium hoffmani* К22, *E. coli* более чем в 14 раз; для *P. aerugenosa*, *C. tropicalis* PY23, *C. krusei* PY22, *C. albicans* NCTC 885-653 примерно в 2 раза; для *C. glabrata* KY22, *C. parapsilosis* KY23 в 1.3 раза. Следовательно, КАМП-1 является перспективной фармацевтическая субстанцией антибактериального и противогрибкового действия.

Таблица 1. Минимально ингибирующие концентрации КАМП-1 на тест-культурах

Инокулюм 10^3	МИК, мкг/мл
S. aureus	6.67 ± 1.67
B. cereus	6.67 ± 1.67
E. coli	8.33 ± 1.67
P. aerugenosa	60.0 ± 20.82
Corynebacterium hoffmani К22	5.56 ± 0.63
C. glabrata KY22	80.0 ± 12.08
C. tropicalis PY23	40.0 ± 4.08
C. krusei PY22	40.0 ± 3.34
C. parapsilosis KY23	80.0 ± 10.22
C. albicans NCTC 885-653	40.0 ± 5.58

При использовании концентрации 7.0 мг/кг не регистрировалась гибель животных. Токсический эффект исследуемого вещества наблюдался в диапазоне доз от 7.5 до 10 мг/кг. В среднем смерть животного наступала в течение первой минуты после введения. Результаты по количеству павших и выживших животных в исследуемых группах представлено в табл. 2.

В концентрациях от 7.5 до 10 мг/кг токсический эффект исследуемой субстанции у каждого животного проявлялся схожей симптоматикой: замедление реакции, учащенное дыхание, отсутствие реакции на звуковые раздражители, отказ от корма и еды, клонические судороги. В данном диапазоне доз у некоторых животных наблюдалось восстановление витальных функций. При воздействии КАМП-1 не было зафиксировано различий в симптомах между самками и

самцами. Результаты по выявленным летальным дозам исследуемой субстанции представлены в табл. 3.

Таблица 2. Результаты исследования мышей после внутривенного введения КАМП-1

Доза, мг/кг	Конвенциональные мыши		Мыши SPF-статуса	
	выжившие	павшие	выжившие	павшие
7.0	5	0	6	0
7.5	5	0	5	1
8.0	4	1	5	1
8.5	4	1	4	2
9.0	3	2	3	3
9.5	2	3	2	4
10.0	0	5	0	6

Таблица 3. Параметры острой токсичности после однократного внутривенного введения КАМП-1 у мышей

Группа животных	LD_{10}, мг/кг	LD_{50}, мг/кг	LD_{90}, мг/кг
Конвенциональные мыши	7.86	9.09	10.32
Мыши SPF-статуса	7.48	8.88	10.28

При вскрытии не были выявлены макроскопические изменения внутренних органов (легкие, сердце, печень, головной мозг, селезенка, кишечник). У исследуемых животных, подвергнутых эвтаназии, не наблюдалось существенных визуальных отличий внутренних органов от интактных мышей. Диапазоны летальных доз и расчетные дозы LD_{10}, LD_{50}, LD_{90} при изучении на животных с наличие микрофлоры и SPF-статуса фактически совпадают. Различия наблюдались лишь в незначительном количестве павших животных, что объясняется разным числом подопытных в обоих группах. Следовательно, наличие микробиома не сказывается на результатах острой токсичности, что в последующем позволит изучить субхроническую, хроническую токсичность в стандартных условиях вивария. Можно предположить, что гибель животных в эксперименте наступала от циркуляторного шока, спровоцированный пептидомиметиком при взаимодействии с рецепторным аппаратом эндотелия.

Выводы: таким образом, КАМП-1 может рассматриваться как перспективная противомикробная фармацевтическая субстанция. Данное обстоятельство подтверждается наличием минимальной ингибирующей концентрацией для широкого спектра аттестованных штаммов, которая не превышает более 100 мг/л. При использовании КАМП-1 наблюдается дозазависимый летальный эффект. При однократном внутривенном введении в диапазоне доз 7.5–10.0 мг/кг регистрируется остротоксическое действие данной субстанции, которое проявлялось гибелью животных, полулетальная доза 8.88 мг/кг, LD_{10} (7.48 мг/кг), LD_{90} (10.28 мг/кг).

Финансирование: Исследование выполнено по теме государственного задания № FWNR-2024-0005.

Analysis of primary pharmacodynamic effects and possible cytotoxic effects of synthetic peptidomimetics SAMP-1

Lyubushkina E.M.*, Bondareva E.A., Soldatova M.S., Madonov P.G.
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
** lizaveta.lyubushkina@list.ru*

Key words: antibiotic resistance; SAMP-1; MIC; acute toxicity

Motivation and Aim: Antibacterial resistance is a threat to public health; according to the estimation of the team led by O'Neil, by 2050, mortality will reach up to 10 million people per year, and financial losses will lead to losses of more than $100 trillion [1]. Modern technologies of antibiotic therapy are implemented at the expense of several groups of antibiotics, to most of which resistance has already been formed. The search for new antibacterial agents to which resistance has not yet formed is an effective solution to this problem. Today, synthetic antimicrobial peptidomimetics are of interest due to the lack of a specific microbial target, and, therefore, lower risk of resistance development to them. The aim of this work was to determine the parameters of antibacterial and antifungal activities, as well as a detailed study of acute toxicity of an innovative peptidomimetic (SAMP-1).
Methods and Algorithms: SAMP-1 consists of an arginine residue, tryptophan modified with tert-butyl groups, and arginine with a phenethylamide group. The molecular mass of the peptide is 788 Dalton. Test cultures of certification strains of bacteria and fungi were used to test antimicrobial activity. The method of serial dilutions of SAMP-1 in liquid nutrient medium was used. A suspension of inoculum of the daily culture was added to a series of twofold dilutions of the substance. After incubation, sowing was performed on a dense nutrient medium with subsequent counting of CFU. The lowest dose of the substance at which complete inhibition of bacterial culture growth was observed was considered the minimum inhibitory concentration (MIC). In each study group, the mean and standard error were calculated based on the results of repeated experiments. The obtained results were processed using Biostatistics statistical software package. In "*in vitro"* studies, the MIC of the studied substance for test strains did not exceed 100 mg/l, which, according to the guidelines for preclinical studies of substances/substances, characterizes it as a promising antimicrobial agent. Acute toxicity parameters were further determined after a single injection into the retroorbital sinus and daily observation of the animals for 14 days. Mongrel mature mice of both sexes were used. The first group of mice were obtained and grown under conventional conditions (n = 40) and the second group of mice was grown under SPF-vivarium conditions (n = 48). The main distinguishing feature was the presence of the microbiome in conventional mice in order to examine differences in outcomes between non-sterile and sterile subjects maintained under SPF-vivarium conditions. The first and second groups included CD-1-line mice of both sexes with external good exterior parameters and weight range from 25–35 g. The animals were kept under standard care conditions with free access to food and water. To study the acute toxicity of the investigated substance, lethal doses LD_{10}, LD_{50}, LD_{90} were calculated using the StatPlus software by the probit analysis method according to V.B. Prozorovsky. SAMP-1 solution was made

in different concentrations from 7.0–10 mg/kg and administered to animals at the rate of 1 µl per 1 gram of weight. The fallen animals were necroscopically examined and after 14 days, the surviving animals were euthanized using carbon dioxide and further macroscopic examination of internal organs was carried out.

Results: The obtained data demonstrate that SAMP-1 has antimicrobial and antifungal activity, which is shown in Table 1. MIC of SAMP-1 is less than the permissible limit of 100 mg/l (100 µg/ml) on the studied certified test-cultures: *S. aureus*, *B. cereus*, *Corynebacterium hoffmani* K22, *E. coli* more than 14 times; for *P. aerugenosa*, *C. tropicalis* PY23, C. krusei PY22, *C. albicans* NCTC 885-653 approximately 2 times; for *C. glabrata* KY22, *C. parapsilosis* KY23 1.3 times. Consequently, SAMP-1 is a promising pharmaceutical substance of antibacterial and antifungal action.

Table 1. Minimally inhibitory concentrations of SAMP-1 on test cultures

Inoculum 10^3	MIC, µg/ml
S. aureus	6.67 ± 1.67
B. cereus	6.67 ± 1.67
E. coli	8.33 ± 1.67
P. aerugenosa	60.0 ± 20.82
Corynebacterium hoffmani K22	5.56 ± 0.63
C. glabrata KY22	80.0 ± 12.08
C. tropicalis PY23	40.0 ± 4.08
C. krusei PY22	40.0 ± 3.34
C. parapsilosis KY23	80.0 ± 10.22
C. albicans NCTC 885-653	40.0 ± 5.58

No death of animals was registered at the concentration of 7.0 mg/kg. Toxic effect of the investigated substance was observed in the dose range from 7.5 to 10 mg/kg. On average, animal death occurred within the first minute after administration. The results of the number of fallen and survived animals in the studied groups are presented in Table 2.

Table 2. Results of mice after intravenous administration of SAMP-1

Dose, mg/kg	Conventional mice		SPF-status mice	
	survived	deceased	survived	deceased
7.0	5	0	6	0
7.5	5	0	5	1
8.0	4	1	5	1
8.5	4	1	4	2
9.0	3	2	3	3
9.5	2	3	2	4
10.0	0	5	0	6

In concentrations from 7.5 to 10 mg/kg toxic effect of the tested substance in each animal was manifested by similar symptoms: slowed reaction, rapid breathing, lack of reaction to sound stimuli, refusal of feed and food, clonic convulsions. In this dose range, some animals showed recovery of vital functions. No differences in symptoms between females and males were recorded during exposure to SAMP-1. The results for the detected lethal doses of the tested substance are presented in Table 3.

Table 3. Parameters of acute toxicity after a single intravenous injection of SAMP-1 in mice

Animal group	LD_{10}, mg/kg	LD_{50}, mg/kg	LD_{90}, mg/kg
Conventional mice	7.86	9.09	10.32
SPF-status mice	7.48	8.88	10.28

No macroscopic changes of internal organs (lungs, heart, liver, brain, spleen, intestine) were detected at autopsy. No significant visual differences of internal organs from intact mice were observed in euthanized animals under study. The ranges of lethal doses and calculated doses LD_{10}, LD_{50}, LD_{90} when studied on animals with the presence of microflora and SPF status actually matched. Differences were observed only in the minor number of dead animals, which is explained by the different number of subjects in both groups. Consequently, the presence of microbiome does not affect the results of acute toxicity, which in the future will allow to study subchronic, chronic toxicity under standard vivarium conditions.

It can be assumed that the death of animals in the experiment was due to circulatory shock provoked by the peptidomimetic in interaction with the receptor apparatus of the endothelium.

Conclusion: Thus, SAMP-1 can be considered as a promising antimicrobial pharmaceutical substance. This fact is confirmed by the presence of the minimum inhibitory concentration for a wide range of certified strains, which does not exceed more than 100 mg/L. A dose-dependent lethal effect is observed when using SAMP-1. At a single intravenous injection in the dose range of 7.5–10.0 mg/kg the acutely toxic effect of this substance is registered, which was manifested by death of animals, semi-lethal dose 8.88 mg/kg, LD_{10} (7.48 mg/kg), LD_{90} (10.28 mg/kg).

Funding: The research was carried out under the subject of the state assignment No. FWNR-2024-0005.

Список литературы/References

1. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations. London, 2016. Available at: https://amr-review.org/ (Accessed May 1, 2024)
2. Миронов А.Н. Руководство по проведению доклинических исследований лекарственных средств. Часть первая. М., 2012
3. Lashin S.A., Matushkin Yu.G. Haploid evolutionary constructor: new features and further challenges. *In Silico Biol.* 2011;11(3-4):125-135

doi 10.18699/bgrs2024-9.4-09

Модифицированный серебром сорбент: исследование влияния на гемостатические реакции «*in vitro*»

Рачковская Л.Н.[1], Момот А.П.[2], Рачковский Э.Э.[1]*, Мамаев А.Н.[2], Федоров Д.В.[3], Смагин А.А.[1], Нимаев В.В.[1], Летягин А.Ю.[1]

[1] *Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия*

[2] *Алтайский филиал ФГБУ «Национальный медицинский исследовательский центр гематологии» Министерства здравоохранения Российской Федерации, Барнаул, Россия*

[3] *Алтайский государственный медицинский университет Министерства здравоохранения Российской Федерации, Барнаул, Россия*

* *reed@academ.org*

Ключевые слова: серебро; оксид алюминия; полидиметилсилоксан; сорбент; тромбоциты; гемокоагуляция; безопасность

Саногенный эффект гемосорбционной детоксикации при различных видах патологии может быть усилен путем модифицирования поверхности сорбента биологически активными компонентами с антибактериальными свойствами. *Цель исследования*: изучить влияние серебросодержащего сорбента на основе пористого оксида алюминия и полидиметилсилоксана на изменение количества тромбоцитов в ходе моделирования гемосорбции и на особенности гемостатического ответа при контакте сорбента с кровью *in vitro*. Исследовали сорбент, модифицированный нанокластерным серебром (Ag, 0,1 %) Al_2O_3@ПДМС-Ag в сравнении с сорбентом без серебра – Al_2O_3@ПДМС. Сорбенты контактировали с донорской кровью при перемешивании с помощью шейкера (99 оборотов/мин) в течение 2 мин (Протокол № 9 от 30.09.2022 ЛЭК). Все исследования (кроме подсчета количества тромбоцитов) проводились с плазмой крови (бедная тромбоцитами плазма – БТП). Оценка ответа системы гемостаза включала в себя: подсчет количества тромбоцитов, а также определение ряда хронометрических и амидолитических показателей. Контакт исследованных сорбентов с кровью вызывает умеренное снижение количества тромбоцитов (на 7–8 %) от их исходного уровня в донорской крови. Отмечена активация свертывания крови по укорочению показателей каолинового времени для обоих сорбентов (26–27 %). Показатель силиконового времени свертывания для сорбента с серебром был близок к показателю исходной крови, для сорбента без серебра наблюдалось укорочение времени на 8 %. В тех же условиях содержание фибриногена активность антитромбина и уровень плазминогена практически не изменились. Полученные результаты свидетельствуют о потенциальной возможности получения новых модифицированных сорбентов с антибактериальными свойствами, имеющих минимальное повреждающее влияние на кровь, и необходимости продолжения исследований в этом направлении.

Финансирование: Исследование поддержано Государственным заданием НИИКЭЛ – филиал ИЦиГ СО РАН (№ FWNR-2022-0009).

Silver-modified sorbent: study of the effect on hemostatic reactions «*in vitro*»

Rachkovskaya L.N.[1], Momot A.P.[2], Rachkovsky E.E.[1]*, Mamaev A.N.[2], Fedorov D.V.[3], Smagin A.A.[1], Nimaev V.V.[1], Letyagin A.Yu.[1]

[1] *Recearch Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia*

[2] *ALTAI branch of the Federal State Budgetary Institution "National Medical Research Center of Hematology" of the Ministry of Health of the Russian Federation, Altai Krai, Barnaul, Russia*

[3] *FSBEI of Higher Education "Altai State Medical University", Barnaul, Russia*

* *reed@academ.org*

Key words: silver; aluminum oxide; polydimethylsiloxane; sorbent; platelets; hemocoagulation; safety

The sanogenic effect of hemosorption detoxification in various types of pathology can be enhanced by modifying the surface of the sorbent with biologically active components with antibacterial properties. The aim of the study was to study the effects of a silver-containing sorbent based on porous aluminum oxide and polydimethylsiloxane on changes in the number of platelets during hemosorption modeling and on the features of the hemostatic response when the sorbent comes into contact with blood in vitro. A sorbent modified with nanocluster silver (Ag, 0.1 %) Al_2O_3@PDMS-Ag was studied in comparison with a sorbent without silver – Al_2O_3@PDMS. The sorbents were in contact with the donor blood while stirring with a shaker (99 revolutions/min) for 2 min (Protocol No. 9 of 30.09.2022 LEK). All studies (except for counting the number of platelets) were conducted with blood plasma (platelet-poor plasma). Studies of the response of the hemostatic system included: assessment of platelet count, chronometric and amidolytic parameters. The assessment of the response of the hemostatic system included: counting the number of platelets, as well as determining a number of chronometric and amidolytic parameters. The contact of the studied sorbents with blood causes a moderate decrease in the number of platelets (by 7–8 %) from their initial level in donated blood. Activation of blood coagulation was noted by shortening the kaolin time for both sorbents (26–27 %). The indicator of silicone clotting time for the sorbent with silver was close to that of the original blood, for the sorbent without silver, a shortening of time by 8 % was observed. Under the same conditions, the fibrinogen content, antithrombin activity and plasminogen levels remained virtually unchanged. The results obtained indicate the potential for obtaining new modified sorbents with antibacterial properties that have minimal damaging effects on the blood, and the need to continue research in this direction.

Funding: The study is supported by the State assignment RICEL – branch of the Instituteof Cytology and Genetics SB RAS (No. FWNR-2022-0009).

doi 10.18699/bgrs2024-9.4-10

Нейромодуляторы азаадамантанового типа: перспективы применения при нейродегенеративных патологиях

Сорокина И.В.*, Питухин М.П., Айдагулова С.В., Мешкова Ю.В.,
Пономарев К.Ю., Толстикова Т.Г., Суслов Е.В., Волчо К.П.
Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия
** sorokina@nioch.nsc.ru*

Ключевые слова: азаадамантаны; нейромодуляторы; болезнь Паркинсона; рассеянный склероз

Мотивация и цель: К нейродегенеративным заболеваниям относится группа болезней с прогрессирующей гибелью и постепенной атрофией нервных клеток в определенных структурах центральной нервной системы. Одним из общих патогенетических механизмов является нарушение функционирования нейротрансмиттерных систем, изменяющее баланс тормозных и возбуждающих нейромедиаторов в нервных тканях, что на уровне целого организма проявляется в возникновении симптомов неврологического дефицита в виде двигательных, когнитивных и вегетативных расстройств. Ведущим патогенным фактором, в данном случае, можно считать повышение активности глутаматной системы и, связанного с ней феномена эксайтотоксичности, возникающего за счет усиления кальциевого тока в нейроны через каналы NMDA-рецепторов. Установлен существенный вклад данного феномена в патогенез болезней Паркинсона, Гентингтона, рассеянного склероза, сенильной деменции, и других нейродегенеративных патологий [1]. Данные факты повышают интерес к разработке современных «антиглутаматных» стратегий и поиску препаратов с нейропротекторным действием, прежде всего среди блокаторов глутаматных рецепторов. Перспективным классом соединений в этой связи являются производные адамантана, на основе которого получены три лекарственных препарата, применяемые при болезнях Альцгеймера (мемантин), Паркинсона (амантадин, гимантан) и рассеянном склерозе (амантадин).

Амантадин (мидантан) используется при лечении болезни Паркинсона (БП), а также в терапии рассеянного склероза, хореи Гентингтона, пост-COVID19 синдрома и др. Амантадин оказывает противовирусное, нейропротекторное, умеренное противовоспалительное действие, а также снижает акинезию, тремор и физическую слабость пациентов за счет модуляции нейротрансмиссии (стимуляция дофаминергических и частичная блокада холинергических нейронов) [1]. Нейропротекторное действие препарата обеспечивается частичной блокадой NMDA глутаматных рецепторов, которые играют важную роль в развитии нейровоспаления и апоптоза клеток нервной ткани [2]. Однакоамантадин имеет ряд побочных эффектов, таких как сонливость, тремор, судороги, а также нарушения со стороны желудочно-кишечного тракта и мочевого пузыря. Кроме того, при длительном курсовом применении амантадин вызывает привыкание.

В НИОХ СО РАН получены три полициклических азотистых аналога амантадина, содержащие атомы азота в узловых местах адамантанового каркаса: 5,7-диметил-1,3-диазаадамантан-6-он (К1-238), 6-амино-5,7-диметил-1,3-диазаадамантан (К1-282) и 7-амино-1,3,5-триазаадамантан (К1-302). Целью исследования являлась

оценка влияния агентов на симптомы дискинезии у мышей CD-1 в моделях лекарственного паркинсонизма, а также на выраженность неврологического дефицита у мышей C57Bl/6 в модели токсической купризоновой демиелинизации.
Методы: Эксперименты проводили согласно [3] на мышах-самцах CD-1 и C57Bl/6, полученных из вивария ФИЦ ФТМ. Исследуемые соединения вводили в дозе 20 мг/кг внутрибрюшинно, контрольной группе вводили физиологический раствор.
Модель галоперидоловой каталепсии. Исследуемые соединения вводили за 15 мин до введения блокатора D2-дофаминовых рецепторов галоперидола в дозе 1 мг/кг. Через 60, 120 и 180 мин животное устанавливали передними лапами на перекладину («поза лектора») и фиксировали продолжительность всех периодов замирания (каталепсии) в течение 2 мин. Дофаминергическую активность соединений оценивали по снижению общей продолжительности всех периодов каталепсии в группе относительно контроля.
Модель ареколинового тремора. Исследуемые агенты вводили за 30 мин до подкожного введения М-холиномиметика ареколина (в дозе 25 мг/кг). Фиксировали продолжительность тремора. Действие агентов в отношении снижения активности М-холинергической системы оценивали по уменьшению продолжительности судорожной активности.
Модель йохимбиновой токсичности. Исследуемые агенты вводили за 30 мин до внутрибрюшинного введения раствора йохимбина, селективного α2-адреноблокатора, в дозе 55 мг/кг. Эффект потенцирования адренергической передачи оценивали по суммарной продолжительности жизни всех мышей в группе и количеству павших животных в течение 30 мин после введения никотина.
Модель рассеянного склероза (купризоновой демиелинизации). Самцам мышей C57BL/6 (кроме интактных) заменяли питьевую воду на 0,3 % раствор купризона. Исследуемые агенты вводили три раза в неделю в течение 40 дней. Оценку моторно-двигательных и поведенческих эффектов агентов в условиях токсической демиелинизации проводили посредством анализа двигательной и исследовательской активности в приборе TruScan («Coulbourn Inst.», США) по следующим показателям: общему времени движения, скорости движения, пройденной дистанции, времени в неподвижном состоянии, а также по числу и времени обследованных отверстий в полу камеры. Показатели регистрировали в течение 2 мин у каждой мыши на 40 день эксперимента.
Результаты: **Модель галоперидоловой каталепсии**. Через 60 мин после введения галоперидола значимого различия между всеми группами не наблюдалось, через 120 мин наблюдали статистически значимое снижение продолжительности каталепсии во всех исследуемых группах по отношению к контролю, а через 180 мин достоверное уменьшение продолжительности сохранялось только в группах К-238 и К-302, причем у группы К-238 также обнаружен значимый эффект по сравнению с АМД. Согласно этим данным, можно сделать вывод о том, что все исследуемые агенты в условиях блокады D2-дофаминовых рецепторов галоперидолом проявили дофаминергическую активность, а агенты К-238 и К-302 сохраняли эту активность дольше чем АМД. Таким образом, установлено, что азаадамантаны способны снижать акинезию и ригидность, характерную для болезни Паркинсона.
Модель ареколинового тремора. На фоне введения исследуемых веществ К-282 и К-302 было обнаружено значимое уменьшение продолжительности тремора,

вызванного ареколином по сравнению с группой АМД и контролем. Данный факт свидетельствует о наличии центральной М-холиноблокирующей активности у данных агентов. Соединения К-282 и К-302 являются потенциально эффективными агентами для предупреждения тремора, вызванного гиперактиваицей М-холинорецепторов хвостатого ядра, аналогичной дрожательной форме болезни Паркинсона.

Модель йохимбиновой токсичности. Йохимбин селективно блокирует α2-адренорецепторы, тем самым вызывая увеличение поступления норадреналина в синаптическую щель и оказывая токсическое действие. Установлено, что агент К-282 уменьшает суммарную продолжительность жизни после введения йохимбина в 1,7 раз, а АМД, К-238 и К-302 – в 1,2–1,4 раза по сравнению с контролем. Гибель всех животных в группе (6 из 6 животных) в течение времени наблюдения была зафиксирована в группах К-238, К-282, К-302; в группе АМД погибло 83 % мышей (5 из 6), а в контрольной группе гибель наступала лишь в 69,2 % случаев (9 из 13). Таким образом, все исследуемые агенты усиливают токсическое действие йохимбина и оказывают потенцирующее действие на адренергическую систему. Сделано предположение о способности агентов блокировать действие фермента моноаминоксидазы в условиях увеличенного высвобождения норадреналина в синаптическую щель.

Модель токсической демиелинизации, вызванной купризоном. На фоне хронической интоксикации купризоном, на 40-й день опыта снижение локомоторных показателей в контрольной группе относительно интактной было статистически незначимым, однако показано, что агент К-238 достоверно превосходил интактную, контрольную и группу АМД по общему времени локомоторной активности, дистанции и скорости движения. Соединение К-302 достоверно повысило общую продолжительность двигательных актов по сравнению с контролем, а также показатели дистанции и скорости движения относительно группы АМД. У агента К-302 также наблюдалась достоверно более высокая исследовательская активность относительно интактной, контрольной и группы АМД. Соединение К-282 не оказало достоверного эффекта на показатели локомоторной и исследовательской активности относительно контроля и АМД. Таким образом, соединения К-238 и К-302 в условиях нейротоксического действия купризона оказывают стимулирующий эффект на двигательную и исследовательскую активность. Такие результаты могут быть связаны с влиянием исследуемых агентов на дофаминергическую и/или адренергическую трансмиссию в нервной системе.

Выводы: Полученные результаты свидетельствуют о перспективности применения нейромодуляторов азаадамантанового типа при нейродегенеративных патологиях

Финансирование: Работа выполнена при финансовой поддержке РНФ, проект 23-25-00428.

Neuromodulators of the azaadamantane type: prospects for use in neurodegenerative pathologies

Sorokina I.V.*, Pitukhin M.P., Aidagulova S.V., Meshkova Yu.V., Ponomarev K.Yu., Tolstikova T.G., Suslov E.V., Volcho K.P.

N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry Russian Academy of Sciences, Novosibirsk, Russia

** sorokina@nioch.nsc.ru*

Key words: azaadamantanes; neuromodulators; Parkinson's disease; multiple sclerosis

Motivation and Aim: Neurodegenerative diseases is a group of diseases with progressive death and gradual atrophy of nerve cells in certain structures of the central nervous system. One of the common pathogenetic mechanisms is disruption of the functioning of neurotransmitter systems, changing the balance of inhibitory and excitatory neurotransmitters in nerve tissues, which at the level of the whole organism manifests itself in the occurrence of symptoms of neurological deficit in the form of motor, cognitive and autonomic disorders. The leading pathogenic factor, in this case, can be considered an increase in the activity of the glutamate system and the associated phenomenon of excitotoxicity, which occurs due to increased calcium current into neurons through NMDA receptor channels. A significant contribution of this phenomenon to the pathogenesis of Parkinson's, Huntington's, multiple sclerosis, senile dementia, and other neurodegenerative pathologies has been established [1]. These facts increase interest in the development of modern "anti-glutamate" strategies and the search for drugs with a neuroprotective effect, primarily among glutamate receptor blockers. A promising class of compounds in this regard are adamantane derivatives, on the basis of which three drugs used for Alzheimer's (memantine), Parkinson's (amantadine, gimantane) and multiple sclerosis (amantadine) diseases have been obtained.

Amantadine (midantan) is used in the treatment of Parkinson's disease (PD), as well as in the treatment of multiple sclerosis, Huntington's chorea, post-COVID19 syndrome, etc. Amantadine has antiviral, neuroprotective, moderate anti-inflammatory effects, and also reduces akinesia, tremor and physical weakness of patients by modulating neurotransmission (stimulation of dopaminergic and partial blockade of cholinergic neurons) [1]. The neuroprotective effect of the drug is ensured by partial blockade of NMDA glutamate receptors, which play an important role in the development of neuroinflammation and apoptosis of nervous tissue cells [2]. However, amantadine has a number of side effects, such as drowsiness, tremor, convulsions, as well as gastrointestinal and bladder disorders. In addition, with long-term course use, amantadine is addictive.

Three polycyclic nitrogenous analogues of amantadine containing nitrogen atoms in the nodes of the adamantane framework were obtained at the Scientific Research Institute of Organic Chemistry of the Siberian Branch of the Russian Academy of Sciences: 5,7-dimethyl-1,3-diazaadamantan-6-one (K1-238), 6-amino-5,7-dimethyl-1,3-diazaadamantane (K1-282) and 7-amino-1,3,5-triazaadamantane (K1-302). The purpose of the study was to evaluate the effect of agents on symptoms of dyskinesia in CD-1 mice in models of drug-induced parkinsonism, as well as on the severity of neurological deficits in C57Bl/6 mice in a model of toxic cuprizone demyelination.

Methods: Experiments were carried out according to [3] on male CD-1 and C57Bl/6 mice obtained from the vivarium of the FRC FTM. The test compounds were administered at a dose of 20 mg/kg intraperitoneally, the control group was administered saline solution.

Model of haloperidol catalepsy. The test compounds were administered 15 min before administration of the D2-dopamine receptor blocker haloperidol at a dose of 1 mg/kg. After 60, 120 and 180 min, the animal was placed with its front paws on the crossbar ("lecturer's pose") and the duration of all periods of freezing (catalepsy) was recorded for 2 min. The dopaminergic activity of the compounds was assessed by reducing the total duration of all periods of catalepsy in the group relative to the control.

Model of arecoline tremor. The study agents were administered 30 min before subcutaneous administration of the M-cholinomimetic arecoline (dose – 25 mg/kg). The duration of tremor was recorded. The effect of the agents in reducing the activity of the M-cholinergic system was assessed by reducing the duration of convulsive activity.

Yohimbine toxicity model. The study agents were administered 30 min before intraperitoneal administration of a solution of yohimbine, a selective α2-blocker, at a dose of 55 mg/kg. The effect of potentiation of adrenergic transmission was assessed by the total life expectancy of all mice in the group and the number of dead animals within 30 min after nicotine administration.

Model of multiple sclerosis (cuprizone demyelination). Male C57BL/6 mice (except intact ones) had their drinking water replaced with a 0.3 % cuprizone solution. Study agents were administered three times per week for 40 days. Evaluation of the motor-motor and behavioral effects of agents in conditions of toxic demyelination was carried out by analyzing motor and exploratory activity in the TruScan device (Coulbourn Inst., USA) according to the following indicators: total movement time, movement speed, distance traveled, time immobile, as well as by the number and time of inspection of holes in the floor of the chamber. The indicators were recorded for 2 min in each mouse on the 40th day of the experiment.

Results: **Model of haloperidol catalepsy.** 60 min after the administration of haloperidol, no significant difference was observed between the groups; after 120 min, a statistically significant decrease in the duration of catalepsy was observed in all study groups relative to the control, and after 180 min, a significant decrease in duration remained only in groups K-238 and K-302, Moreover, a significant effect was also found in the K-238 group compared to AMD. According to these data, it can be concluded that all the studied agents showed dopaminergic activity under conditions of blockade of D2-dopamine receptors with haloperidol, and agents K-238 and K-302 retained this activity longer than AMD. Thus, it has been established that azaadamantanes are able to reduce akinesia and rigidity characteristic of Parkinson's disease.

Model of arecoline tremor. Upon administration of the test substances K-282 and K-302, a significant decrease in the duration of tremor caused by arecoline was found compared to the AMD group and control. This fact indicates the presence of central M-anticholinergic activity in these agents. Compounds K-282 and K-302 are potentially effective agents for the prevention of tremor caused by hyperactivation of M-cholinergic receptors in the caudate nucleus, similar to the tremor form of Parkinson's disease.

Yohimbine toxicity model. Yohimbine selectively blocks α2-adrenergic receptors, thereby causing an increase in the flow of norepinephrine into the synaptic cleft and causing a toxic effect. It was found that the agent K-282 reduces the total life expectancy after administration of yohimbine by 1.7 times, and AMD, K-238 and K-302 – by 1.2–

1.4 times compared to the control. The death of all animals in the group (6 out of 6 animals) during the observation period was recorded in groups K-238, K-282, K-302; in the AMD group, 83 % of mice died (5 out of 6), and in the control group, death occurred only in 69.2 % of cases (9 out of 13). Thus, all studied agents enhance the toxic effect of yohimbine and have a potentiating effect on the adrenergic system. An assumption has been made about the ability of the agents to block the action of the enzyme monoamine oxidase under conditions of increased release of norepinephrine into the synaptic cleft.

Model of toxic demyelination caused by cuprizone. Against the background of chronic intoxication with cuprizone, on the 40th day of the experiment, the decrease in locomotor indicators in the control group relative to the intact group was statistically insignificant, however, it was shown that the K-238 agent was significantly superior to the intact, control and AMD groups in the total time of locomotor activity, distance and speed of movement. Compound K-302 significantly increased the total duration of motor acts compared to the control, as well as indicators of distance and speed of movement relative to the AMD group. Agent K-302 also showed significantly higher exploratory activity relative to the intact, control, and AMD groups. Compound K-282 did not have a significant effect on indicators of locomotor and exploratory activity relative to control and AMD. Thus, compounds K-238 and K-302, under the conditions of the neurotoxic effect of cuprizone, have a stimulating effect on motor and exploratory activity. Such results may be related to the effect of the studied agents on dopaminergic and/or adrenergic transmission in the nervous system.

Conclusion: The results obtained show the prospects of using azaadamantane-type neuromodulators in neurodegenerative pathologies.

Funding: The work was carried out with the financial support of the Russian Science Foundation, project 23-25-00428.

Список литературы/References

1. Wanka L., Iqbal K., Schreiner P.R. The lipophilic bullet hits the targets: medicinal chemistry of adamantane derivatives. *Chem Rev*. 2013;113(5):3516-3604
2. Dong X., Wang Y., Qin Z. Molecular mechanisms of excitotoxicity and their relevance to pathogenesis of neurodegenerative diseases. *Acta Pharmacol Sin*. 2009;30(4):379-387
3. Миронов А.Н., Бунятян Н.Д. Руководство по проведению доклинических исследований лекарственных средств. Часть первая. М.: Гриф и К., 2012

doi 10.18699/bgrs2024-9.4-11

Применение растительных экстрактов для коррекции нарушений метаболических процессов

Толстикова Т.Г.*, Киселёва Д.А., Аньков С.В., Жукова Н.А.
Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия
** tg_tolstikova@mail.ru*

Ключевые слова: фитотерапия; левзея; урсоловая кислота; экдистен; коррекция; мыши; крысы; метаболический процесс

Мотивация и цель: Фитотерапия, являясь методологической основой восстановительной медицины, применяется для профилактики и лечения нарушений обмена веществ, повышения сопротивляемости организма к различным стрессовым условиям (обеспечивая тем самым развитие адаптации и гомеостаз организма) [1]. Биологическая активность лекарственных растений и растительных экстрактов объясняется наличием в их составе вторичных метаболитов (полифенолов, тритерпенов, флавоноидов, гликозидов и т. д.) [2, 3]. Современные исследования доказали высокую эффективность комплексов биологически активных веществ растений (в том числе в виде экстрактов) для коррекции нарушений метаболических процессов, в том числе физического и биологического характера. Кроме того, разработка комплексных растительных композиций может обеспечить наличие синергических эффектов у конечного продукта и более широкий спектр проявляемых фармакологических свойств. Перспективными в этом плане остаются малоизученные аспекты сочетанности действия двух вторичных метаболитов: пентациклических тритерпеноидов (урсоловой кислоты) и фитоэкдистероидов (экдистена). В связи с чем, основной целью исследования было изучение влияния композиции на основе экстрактов левзеи и клюквы, содержащих экдистен и урсоловую кислоту соответственно, на метаболические процессы.
Методы и алгоритмы: Работа выполнена на 100 мышах линии C57BL/6J (70 самцах и 30 самках), 424 аутбредных мышах CD-1 (272 самцах и 152 самках), 14 самцах мышей линии C57BL/6^{Ay}, 80 крысах сток Вистар (24 самцах и 56 самках), полученных из вивария ФИЦ Института цитологии и генетики СО РАН. Содержание животных и дизайн экспериментального исследования были одобрены этическим комитетом НИОХ СО РАН (протокол от 28.08.2021; № 38 от 15.12.2021; № 150 от 13.05.2022; № 431 от 18.05.2023). Все манипуляции с животными проводились в соответствии с законодательством РФ, Решением «Об утверждении Правил надлежащей лабораторной практики Евразийского экономического союза в сфере обращения лекарственных средств» от 03.11.2016 № 81 и положениями Директивы 2010/63/EU Парламента ЕС и Совета Европейского Союза от 22.09.2010 о защите животных, используемых в научных целях. Исследуемые соединения: метилтретбутилового эфира экстракт шрота клюквы, содержащий 40 % урсоловой кислоты, стандартизированный по урсоловой кислоте (Инжиниринговый центр НИОХ СО РАН, г. Новосибирск); водно-спиртовой экстракт корней левзеи сафлоровидной, содержащий 0.31 % экдистена, стандартизированный по экдистену (20-гидроксиэкдизону) («ООО

Экстракты Алтая», г. Барнаул); композиция экстрактов левзеи (0.31 % экдистена) и шрота клюквы (40 % урсоловой кислоты); экдистен, 98 % чистоты (Tokyo Chemical Industry Co., LTD, Japan); урсоловая кислота, 92 % чистоты (Инжиниринговый центр НИОХ СО РАН, г. Новосибирск). Скрининг гипогликемической активности композиции экстрактов левзеи и шрота клюквы в дозах 170:50, 35:250, 70:500 мг/кг при длительном введении (4 недели) как без, так и на фоне физической нагрузки (тест «бег на тредмиле», «принудительное плавание») проводили по показателям динамики массы тела, уровня глюкозы крови, глюкозотолерантного теста [4, 5]. Изучение наличия адаптогенных свойств у подобранного эффективного соотношения композиции (70:500 мг/кг) оценивали в тесте «принудительное плавание». Исследование наличия у композиции анаболической активности проводили на модели изолированной перегрузки скелетной мышцы голени крыс с операцией тенотомии [6]. Влияние композиции в дозе 70:500 мг/кг на уровень тестостерона крови проводили с помощью ИФА анализа. Безопасность и переносимость длительного введения композиции в дозе 70:500 мг/кг и ее отдельных компонентов проводили согласно Руководству по экспериментальному (доклиническому) изучению новых фармакологических веществ [7]. Влияние композиции в дозе 70:500 мг/кг на показатели липидного, углеводного обмена на фоне патологии изучали на диет-индуцированной модели сахарного диабета II типа [8], генетической модели сахарного диабета II типа [9].
Результаты: На основании скрининга выявлено, что композиция экстрактов левзеи и шрота клюквы в дозе 70:500 мг/кг (210 мкг/кг экдистена и 200 мг/кг урсоловой кислоты соответственно) оказывает положительное влияние на показатели углеводного обмена, а именно проявляла гипогликемический эффект на фоне глюкозной нагрузки как без, так и на фоне физической нагрузки. Разработанная композиция в дозе 70:500 мг/кг проявляла аналогичные со смесью индивидуальных действующих веществ фармакологические эффекты в отношении показателей углеводного, белкового, липидного обмена на фоне физической нагрузки (тест «принудительное плавание»), а именно: было отмечено значимое ($p < 0,05$) снижение концентрации глюкозы, лактата, креатинкиназы, холестерина, липопротеинов высокой плотности, увеличение концентрации общего белка, что может являться показателями развития реакций адаптации организма [10]. На модели изолированной перегрузки скелетной мышцы голени крыс композиция в дозе 70:500 мг/кг проявляла анаболический эффект у крыс обоих полов. Полученные значения процента прироста мышцы *m. soleus* на фоне введения композиции коррелировали с показателями смеси индивидуальных действующих веществ, что подтверждает их роль в проявлении композицией анаболического эффекта. При однократном введении композиции в дозе 70:500 мг/кг, экстракта шрота клюквы в дозе 500 мг/кг происходило достоверное увеличение уровня тестостерона у самцов мышей CD-1. В то время как на фоне однократного введения экстракта левзеи в дозе 70 мг/кг не происходило статистически значимого изменения уровня тестостерона. При оценке безопасности и переносимости длительного введения композиции не наблюдалось визуальных признаков интоксикации, гибели животных, токсических эффектов в отношении внутренних органов, биохимических показателей крови. Кроме того, введение композиции в дозе 70:500 мг/кг на фоне нарушенной толерантности к глюкозе (диет-индуцированная модель сахарного диабета II типа, генетической модели сахарного диабета II типа) не усугубляло развитие патологических

процессов в органах-мишенях (печень, поджелудочная железа), оказывало положительное влияние на показатели липидного обмена и способствовало уменьшению размеров адипоцитов белой и бурой жировой ткани, предотвращая развитие гипертрофического ожирения.
Выводы: Композиция экстрактов левзеи и шрота клюквы в дозе 70:500 мг/кг обладает адаптогенными свойствами на фоне физической нагрузки и выраженной анаболической активностью, сопоставимой с активностью смеси действующих веществ – экдистена и урсоловой кислоты. На фоне однократного введения композиции увеличение уровня тестостерона обусловлено наличием в составе композиции экстракта шрота клюквы. Длительное введение композиции не оказывает токсического влияния на организм экспериментальных животных. На фоне врожденной и приобретенной патологии композиция оказывает положительное влияние на показатели липидного обмена и перераспределение жировой ткани.
Финансирование: Исследование выполнено в рамках базовых проектов ФНИ НИОХ СО РАН (№ 122040400038-4).

Plant extracts application for correction disorders of metabolic processes

Tolstikova T.G.*, Kiseleva D.A., An'kov S.V., Zhukova N.A.
Vorozhtzov Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia
** tg_tolstikova@mail.ru*

Key words: phytotherapy; leuzea; ursolic acid; ecdystene; correction; mice; rats; metabolic process

Motivation and Aim: Phytotherapy, being the methodological basis of regenerative medicine, is used for the prevention and treatment of metabolic disorders, increasing the body's resistance to various stressful conditions (thus providing adaptation and body homeostasis) [1]. The biological activity of medicinal plants and plant extracts is explained by the presence of secondary metabolites (polyphenols, triterpenes, flavonoids, glycosides, etc.) [2, 3]. Modern research has proven high efficiency of plant complexes in the form of extracts for correction of metabolic processes disorders, including physical and biological nature. In addition, the complex herbal compositions development may provide synergistic effects in the final product and a wider range of pharmacological properties exhibited. The poorly studied aspects of the combined action of two secondary metabolites: pentacyclic triterpenoids (ursolic acid) and phytoecdysteroids (ecdystene) remain promising. Therefore, the main aim of the study was to investigate the effect of leuzea and cranberry extracts composition, containing ecdystene and ursolic acid, respectively, on metabolic processes.
Methods and Algorithms: The experiments were performed on 100 C57BL/6J mice of both sexes (70 males and 30 females), 424 CD-1 mice of both sexes (272 males and 152 females), 14 male C57BL/6^{Ay} mice and 80 stock Wistar rats (24 males and 56 females). Animals were obtained from the vivarium of the Institute of Cytology and Genetics SB RAS. The protocol of the animal experiment was approved by the Ethics Committee of N.N. Vorozhtsov Institute of Organic Chemistry SB RAS (protocol of 28.08.2021; No. 38 of 15.12.2021; No. 150 of 13.05.2022; No. 431 of 18.05.2023). All manipulations with animals were conducted in strict accordance with the laws and

regulations of the Russian Federation, the Decision "On Approval of the Rules of Good Laboratory Practice of the Eurasian Economic Union in the field of circulation of medicines" No. 81 of 3 November 2016, and the provisions of Directive 2010/63/EU of the European Parliament and of the Council of the European Union of 22 September 2010 on the protection of animals used for scientific purposes. Materials: methyl tert-butyl ether cranberry meal extract containing 40 % ursolic acid, standardized for ursolic acid (Engineering Center of NIOCH SB RAS, Novosibirsk); aqueous-alcoholic radix leuzea extract containing 0.31 % ecdystene, standardized by 20-hydroxyecdysone (Altai Extracts Ltd., Barnaul); composition of leuzea (0.31 % ecdystene) and cranberry meal (40 % ursolic acid) extracts; ecdystene, 98 % purity (Tokyo Chemical Industry Co., LTD, Japan); ursolic acid, 92 % purity (Engineering Center of NIOCH SB RAS, Novosibirsk). Hypoglycemic activity screening of the composition of leuzea and cranberry meal extracts in doses of 170:50, 35:250, 70:500 mg/kg at long-term administration (4 weeks) with and without physical load ("treadmill test", "forced swim test" (FST)) was carried out according to the body weight dynamics, blood glucose level, glucose tolerance test [4, 5]. The adaptogenic properties in the selected composition (70:500 mg/kg) was evaluated in the FST. The anabolic activity of the composition was investigated on a model of isolated overload of the rat skeletal muscle with tenotomy operation [6]. The effect of the composition at a dose of 70:500 mg/kg on the blood testosterone level was carried out by ELISA analysis. Safety and tolerability of long-term administration of the composition and its individual components were carried out according to the Guidelines for experimental preclinical studies of new pharmacological substances [7]. The effect of the composition at a dose of 70:500 mg/kg on lipid and carbohydrate metabolism of experimental animals was studied on a diet-induced model of type II diabetes mellitus [8], genetic model of type II diabetes mellitus [9].

Results: According to the screening results, the composition of leuzea and cranberry meal extracts at a dose of 70:500 mg/kg (210 μg/kg ecdystene and 200 mg/kg ursolic acid respectively) has a positive effect on carbohydrate metabolism parameters (showed hypoglycemic effect on the background of glucose load, with and without physical load). The developed composition at a dose of 70:500 mg/kg showed similar pharmacological effects to the individual active substances in terms of carbohydrate, protein, lipid metabolism in the FST, namely: there was a significant (III < 0.05) decrease of glucose, lactate, creatine kinase, cholesterol, high-density lipoproteins concentration, an increase in total protein concentration, which may be indicators of the development of the organism adaptation reactions [10]. On a model of isolated overload of the rat skeletal muscle, the composition at a dose of 70:500 mg/kg showed anabolic effect in animals of both sexes. The obtained values of *m. soleus* muscle growth at administration of the composition correlated with the indices of individual active substances mixture, which confirms their role in the manifestation of anabolic effect by the composition. A single administration of the composition at a dose of 70:500 mg/kg and cranberry meal extract at a dose of 500 mg/kg resulted in a significant increase in testosterone levels CD-1 male mice. However, there was no statistically significant change in testosterone levels after a single injection of leuzea extract at a dose of 70 mg/kg. When assessing the safety and tolerability of long-term administration of the composition, no visual signs of intoxication, animals died, toxic effects to internal organs, blood biochemical indices were observed. In addition, administration of the composition at a dose of 70:500 mg/kg at impaired glucose tolerance (diet-induced model of type II diabetes mellitus, genetic model of type II diabetes mellitus) did not worsen the course of pathological processes

in target organs (liver, pancreas), had a positive effect on lipid metabolism parameters and promoted the reduction of white and brown fat adipocytes size, preventing the development of hypertrophic obesity.

Conclusion: The composition of leuzea and cranberry meal extracts at a dose of 70:500 mg/kg has adaptogenic properties and pronounced anabolic activity, comparable to the activity of active substances – ecdystene and ursolic acid. The increase in testosterone level at a single administration of the composition is due to the presence of cranberry meal extract in the composition. Long-term administration of the composition has no toxic effect on the organism of experimental animals. The composition has a positive effect on lipid metabolism indicators and redistribution of adipose tissue against the background of pathological changes.

Funding: The study was carried out within the framework of basic projects of FSR NIOCH SB RAS (No. 122040400038-4).

Список литературы/References

1. Colalto C. What phytotherapy needs: Evidence-based guidelines for better clinical practice. *Phytother Res.* 2018;32(3):413-425. doi 10.1002/ptr.5977
2. Bujor O.C., Tanase C., Popa M.E. Phenolic Antioxidants in Aerial Parts of Wild Vaccinium Species: Towards Pharmaceutical and Biological Properties. *Antioxidants (Basel).* 2019;8(12):649. doi 10.3390/antiox8120649
3. Majhi S. Discovery, Development and Design of Anthocyanins-Inspired Anticancer Agents: A Comprehensive Review. *Anticancer Agents Med Chem.* 2022;22(19):3219-3238. doi 10.2174/1871520621666211015142310
4. Odriozola C.P. et al. Reliability of blood lactate as a measure of exercise intensity in different strains of mice during forced treadmill running. *PLoS One*. 2019;14(5):e0215584. doi 10.1371/journal.pone.0215584
5. Ковалева М.А. и др. Применение теста «принудительное плавание» при проведении доклинических исследований. *Международный вестник ветеринарии*. 2015;4:90-95
   [Kovaleva M.A. et al. Application of «forced swimming» test for preclinical trials. *International bulletin of veterinary medicine.* 2015;4:90-95 (in Russian)]
6. Гаджиева Д.М. и др. Сравнительное изучение анаболизирующей активности апилака и метандростенолона на модели изолированной перегрузки скелетной мышцы голени крыс. *Экспериментальная и клиническая фармакология*. 2002;65(1):56-57
   [Gadzhieva D.M. et al. Comparative study of anabolizing activity of apilac and methandrostenolone on a model of isolated overload of the rat skeletal muscle. *Eksp Klin Farmakol*. 2002;65(1):56-57 (in Russian)]
7. Миронов А.Н., Бунатян Н.Д. Руководство по проведению доклинических исследований лекарственных средств. Часть первая. М.: Гриф и К., 2012
   [Mironov A.N., Bunatyan N.D. Manual for conducting preclinical studies of medicinal products. Part one. Moscow: Grif and K., 2012 (in Russian)]
8. Surwit R.S. et al. Diet-Induced Type II Diabetes in C57BL/6J Mice. *Diabetes*. 1988;37(9):1163-1167. doi 10.2337/diab.37.9.1163
9. Derkach K.V. et al. Metabolic parameters and functional state of hypothalamic signaling systems in AY/a mice with genetic predisposition to obesity and the effect of metformin. *Dokl Biochem Biophys*. 2017;477(1):377-381. doi 10.1134/S1607672917060096
10. Mul J.D. et al. Exercise and Regulation of Carbohydrate Metabolism. *Prog Mol Biol Transl Sci*. 2015;135:17. doi 10.1016/bs.pmbts.2015.07.020

doi 10.18699/bgrs2024-9.4-12

Влияние пегелированной гиалуронидазы на ультраструктурную организацию гепатоцитов человека с позиции доклинической оценки безопасности *in vitro*

Швецова А.[1, 3]*, Ершов К.[1, 2], Королев М.[1], Чурин А.[3], Бондаренко Н.[1], Бгатова Н.[1], Солдатова М.[1], Мадонов П.[1, 2]

[1] *Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный медицинский университет МЗ РФ, Новосибирск, Россия*

[3] *Научно-исследовательский институт фармакологии и регенеративной медицины имени Е.Д. Гольдберга ТНИМЦ РАН, Томск, Россия*

* *aleksa-2904@mail.ru*

Ключевые слова: пегелированная гиалуронидаза; ультраструктура; клеточная культура; токсичность

Мотивация и цель: Лекарственные препараты на основе тестикулярной гиалуронидазы активно используются в различных областях медицины. Создан прототип лекарственного средства на основе гиалуронидазы, который предполагает применение *per os* [1]. Разрабатываемое новое лекарственное средство должно быть протестировано в доклинических исследованиях, поскольку изучение токсичности и безопасности является одним из необходимых и ключевых этапов. Основополагающий механизм токсического действия ксенобиотиков заключается в повреждении клеток, которое проявляется формированием нарушений в их структуре или функциональном компоненте, что приводит к гибели [2]. Концепция использования тест-системы культуры клеток *in vitro* для оценки токсичности и фармакологической безопасности лекарственных средств в настоящее время весьма актуальна [3]. Клеточный уровень организации является первым, на котором можно провести анализ нарушений в системе «взаимодействие-структура» субклеточных компонентов, определить механизм повреждающего воздействия [4]. Процесс изучения токсического влияния нового разрабатываемого лекарственного средства на культуре гепатоцитов человека обусловлен, в первую очередь, их лекарственной гепатотоксичностью, являющейся причиной отзыва некоторых лекарственных средств из оборота на фармацевтическом рынке [5]. Безусловно, основной моделью являются эксперименты *in vivo* на животных. Однако применение альтернативных моделей клеточных культур человека популярны и позволяют получить данные, которые в последующем способствуют уменьшению числа экспериментов *in vivo* на лабораторных животных, и которые при экстраполяции на организм Homo sapiens более релевантны. Метод электронной микроскопии позволяет наглядно представить внутриклеточные структурные компоненты. Проанализировав результаты, можно получить общую картину влияния изучаемого средства на цитоархитектонику и, следовательно, его токсическое воздействие.

Целью было провести анализ ультраструктуры культуры клеток гепатоцитов человека при воздействии пегелированной гиалуронидазы с позиции доклинической оценки безопасности и установить ассоциацию взаимосвязь доза-эффект.

Методы и алгоритмы: Объектом исследования является пегелированная тестикулярная гиалуронидаза Н20(ПЭГ-ГИАЛ), подвергнутая электронно-лучевой иммобилизации на полиэтиленоксиде Макрогол 1500 потоком ускоренных электроном (доза 1,5 Мрад) на импульсном линейном ускорителе-10 (г. Новосибирск, Россия). ПЭГ-ГИАЛ представляет собой лиофилизированный светло-серого цвета порошок (ферментативная активность 2800 ЕД в 1 г сухого вещества). Исследование ПЭГ-ГИАЛ проводили в дозах 37, 75 и 150 ЕД/мл.

Клеточные линии. Перевиваемая культура клеток печени человека Chanq liver, получены из коллекции клеток ФБУН ГНЦ ВиБ «Вектор» Роспотребнадзора, Россия. Клетки культивировали на питательной среде RPMI-1640 (Биолот, Россия) с добавлением 10 % эмбриональной телячьей сыворотки (ЭТС, HyClone, US Origin), 40 мкг/мл гентамицина сульфата (Дальхимфарм, Российская Федерация) и 2 мМоль L-глутамина (ICN, США) при температуре 37 °С в CO_2 инкубаторе.

Оценку ультраструктуры культуры клеток гепатоцитов человека при воздействии ПЭГ-ГИАЛ проводили методом электронной микроскопии.

Клетки Chanq liver культивировали в течение 24 ч. Удаляли среду, добавляя при этом полную среду с 1 % содержанием ЭТС, а также добавляли ПЭГ-ГИАЛ в дозах 37, 75 и 150 ЕД/мл. Продолжали культивирование клеток в течение 72 ч в CO_2-инкубаторе (температура 37 °С). Культуральную среду удаляли. Фиксирование клеток осуществляли в 1 % растворе OsO_4 на фосфатном буфере (при pH 7,4). После фиксации проводили процесс дегидратации с помощью серии растворов этилового спирта возрастающей концентрации, далее заключали в эпон (Serva, Германия). На ультратоме Leica UC7/FC7 (Германия/Швейцария) получены полутонкие срезы толщиной 1. Срезы окрашены толуидиновым синим. На световом микроскопе (LEICA DME) отобраны клетки для дальнейшего исследования. Получены ультратонкие срезы толщиной 70–100 нм. Срезы собраны на медные сетки для электронной микроскопии, с помощью насыщенного водного раствора уранилацетата и цитрата свинца отконтрастированы. Полученные образцы исследованы на электронном микроскопе JEM 1400 (Япония) в ЦКП микроскопического анализа биологических объектов СО РАН (г. Новосибирск, Россия).

Результаты: В экспериментах по изучению воздействия ПЭГ-ГИАЛ на ультраструктурную организацию гепатоцитов показано, что во всех изучаемых дозировках наблюдаются некоторые изменения в структуре клеток. При электронно-микроскопическом исследовании гепатоцитов, при культивировании с ПЭГ-ГИАЛ во всех изучаемых дозах, наблюдали значительное возрастание содержания и размеров микроворсинок на мембране клеток, что можно трактовать как увеличение метаболической активности. Цитоплазматическая мембрана не имела нарушений целостности и фрагментации. В условиях культивирования с ПЭГ-ГИАЛ не наблюдалось уменьшения насыщенности клеток органеллами. ПЭГ-ГИАЛ во всех дозах приводит к увеличению содержания структур белкового синтеза клеток – свободных полисомальных комплексов рибосом или мембран гранулярного эндоплазматического ретикулума. Отмечали изменение в содержании аутофагических структур – аутолизосом и аутофагосом,

везикулярных структур. Структура ядра без изменений. Митохондрии имели различную форму и размеры. Так, в структуре митохондрий гепатоцитов, которые культивировали с ПЭГ-ГИАЛ в дозе 37 ЕД/мл наблюдали возрастание содержания крист и увеличение электронной плотности матрикса. В то время как при добавлении ПЭГ-ГИАЛ в дозе 75 ЕД/мл митохондрии располагались группами в клетке и в структуре имели четко очерченные границы мембран со слабо визуализированными кристами. При добавлении максимальной изучаемой дозы ПЭГ-ГИАЛ 150 ЕД/мл в гепатоцитах отмечали гетерогенность в структуре митохондрий: имели различные размеры, форму, электронную плотность и содержание крист. Признаков альтерации и деструкционных процессов митохондрий [6], которые проявляются в стойком набухании и вакуолизации, приводящие к разрывам и дестабилизации мембран, просветлением матрикса и гомогенизацией крист не наблюдали. Изменения в структуре гепатоцитов при дистрофических изменениях отражают как повреждение и нарушение их функций, так и повышенную метаболическую активность и интенсивную функцию. В исследовании [7] показано, что ведущим ультраструктурным изменением клетки является редукция белоксинтезирующих органелл, деструкция митохондриального компартмента и усиление процессов аутофагии.

Выводы: Проведенное экспериментальное исследование продемонстрировало отсутствие токсического влияния пегелированной гиалуронидазы на ультраструктурном уровне гепатоцитов во всех изучаемых дозах. Представленные в работе результаты оценки цитотоксичности ПЭГ-ГИАЛ демонстрируют, что не было выявлено повреждений клеток и нарушения в структуре компонентов. Результаты показали, что происходит активация обменных процессов в клетках, развитие процессов аутофагии, который по мнению авторов [8] является одним из основных процессов поддержания гомеостаза. Активация данного процесса, возможно, происходит в ответ на изменения метаболизма, защите клетки от образования цитотоксических компонентов, выступающий как один из механизмов адаптивного клеточного ответа.

Финансирование: Исследование проведено в рамках государственного задания НИИКЭЛ – филиал ФИЦ ИЦиГ СО РАН по теме № FWNR-2024-0005.

Effect of pegylated hyaluronidase on the ultrastructural organisation of human hepatocytes from the perspective of preclinical in vitro safety assessment

Shvetsova A.[1, 3]*, Ershov K.[1, 2], Korolev M.[1], Churin A.[3], Bondarenko N.[1], Bgatova N.[1], Soldatova M.[1], Madonov P.[1, 2]

[1] *Recearch Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State Medical University, Novosibirsk, Russia*

[3] *Goldberg Research Institute of Pharmacology and Regenerative Medicine, Tomsk, Russia*

* *aleksa-2904@mail.ru*

Key words: Pegelated hyaluronidase; ultrastructure; cell culture; toxicity

Motivation and Aim: Medicinal preparations based on testicular hyaluronidase are actively used in various fields of medicine. A prototype of a hyaluronidase-based drug has been created, which involves per os use [1]. The new drug under development should be tested in preclinical studies, as the study of toxicity and safety is one of the necessary and key stages. The underlying mechanism of toxic action of xenobiotics is cell damage, which is manifested by the formation of disturbances in their structure or functional component, leading to death [2]. The concept of using an *in vitro* cell culture test system to assess the toxicity and pharmacological safety of drugs is currently very relevant [3]. The cellular level of organisation is the first level at which it is possible to analyse disorders in the system «interaction-structure» of subcellular components, to determine the mechanism of damaging effects [4]. The process of studying the toxic effect of a new drug under development on the culture of human hepatocytes is primarily due to their drug-induced hepatotoxicity, which is the reason for the recall of some drugs from circulation in the pharmaceutical market [5]. Undoubtedly, the primary model is *in vivo* animal experiments. However, the use of alternative models of human cell cultures is popular and allows obtaining data that subsequently contribute to reducing the number of *in vivo* experiments on laboratory animals and that are more relevant when extrapolated to the organism of Homo sapiens. The electron microscopy method allows visualisation of intracellular structural components. By analysing the results, it is possible to obtain an overall picture of the effect of the studied agent on cytoarchitectonics and, consequently, its toxic effects.

The aim was to analyse the ultrastructure of human hepatocyte cell culture upon exposure to pegylated hyaluronidase from the perspective of preclinical safety assessment and to establish the association of dose-effect relationship.

Methods and Algorithms: The object of the study is pegylated testicular hyaluronidase H20(PEG-GIAL) subjected to electron-beam immobilisation on polyethylene oxide Macrogol 1500 by a stream of accelerated electrons (dose 1,5 Mrad) on a pulsed linear accelerator-10 (Novosibirsk, Russia). PEG-GIAL is a lyophilised light grey powder (enzymatic activity 2800ED in 1 g of dry substance). PEG-GIAL was studied at doses of 37, 75 and 150 ED/ml.

Cell lines. Transviable culture of human liver Chanq liver cells, obtained from the cell collection of the Federal Biological Research Centre Vector, Russia. Cells were cultured on RPMI-1640 medium (Biolot, Russia) supplemented with 10 % fetal calf serum (ETC, NuClone, US Origin), 40 μg/ml gentamicin sulfate (Dalchimpharm, Russian Federation) and 2 mMol L-glutamine (ICN, USA) at 37 °C in CO_2 incubator.

The ultrastructure of human hepatocyte cell culture under PEG-GIAL exposure was evaluated by electron microscopy.

Chanq liver cells were cultured for 24 hours. The medium was removed while adding complete medium containing 1 % EFV, and PEG-GIAL was added at doses of 37 , 75 and 150 ED/ml. Cell culture was continued for 72 hours in a CO_2 incubator (temperature 37 °C). The culture medium was removed. Cells were fixed in 1 % OsO4 solution on phosphate buffer (at pH 7,4). After fixation, dehydration process was carried out using a series of ethyl alcohol solutions of increasing concentration, then encapsulated in epon (Serva, Germany). Semi-thin sections of thickness 1 were obtained on a Leica UC7/FC7 ultratome (Germany/Switzerland). The sections were stained with toluidine blue. Cells were selected for further study on a light microscope (LEICA DME). Ultra-thin slices with a thickness of 70-100 nm were obtained. The slices were collected on copper grids for electron microscopy and contrasted with saturated aqueous solution of uranyl acetate

and lead citrate. The obtained samples were examined on the electron microscope JEM 1400 (Japan) in the Centre for Microscopic Analysis of Biological Objects of the Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia).

Results: In experiments on studying the effect of PEG-GIAL on the ultrastructural organisation of hepatocytes it was shown that in all studied doses some changes in cell structure were observed. At electron-microscopic study of hepatocytes, when cultured with PEG-GIAL in all studied doses, a significant increase in the content and size of microvilli on the cell membrane was observed, which can be interpreted as an increase in metabolic activity. The cytoplasmic membrane had no disturbances of integrity and fragmentation. Under conditions of cultivation with PEG-GIAL, no decrease in the saturation of cells with organelles was observed. PEG-GIAL at all doses leads to an increase in the content of cell protein synthesis structures – free polysomal complexes of ribosomes or membranes of granular endoplasmic reticulum. Changes in the content of autophagic structures – autolysosomes and autophagosomes, vesicular structures were noted. The structure of the nucleus was unchanged. Mitochondria had different shape and sizes. Thus, in the structure of mitochondria of hepatocytes, which were cultured with PEG-GIAL at a dose of 37 ED/ml, an increase in the content of cristae and an increase in the electron density of the matrix were observed. Whereas, when PEG-GIAL was added at a dose of 75 ED/ml, mitochondria were arranged in groups in the cell and in the structure had well-defined membrane boundaries with poorly visualised cristae. When the maximum studied dose of PEG-GIAL 150ED/ml was added to hepatocytes, heterogeneity in the structure of mitochondria was observed: they had different sizes, shape, electron density and cristae content. No signs of alteration and destructive processes of mitochondria [6], which are manifested in persistent swelling and vacuolisation, leading to ruptures and destabilisation of membranes, matrix luminescence and homogenisation of cristae were observed. Changes in the structure of hepatocytes at dystrophic changes reflect both damage and disturbance of their functions, and increased metabolic activity and intensive function. The study [7] shows that the leading ultrastructural change of the cell is the reduction of protein synthesising organelles, destruction of the mito-chondrial compartment and enhancement of autophagy processes.

Conclusion: The conducted experimental study demonstrated the absence of toxic effect of pegylated hyaluronidase on the ultrastructural level of hepatocytes at all studied doses. The results of PEG-GIAL cytotoxicity evaluation presented in the work demonstrate that no cell damage and disruption in the structure of components were detected. The results showed that there is an activation of metabolic processes in cells, development of autophagy processes, which according to the authors [8] is one of the main processes of homeostasis maintenance. Activation of this process probably occurs in response to changes in metabolism, protecting the cell from the formation of cytotoxic components, acting as one of the mechanisms of adaptive cellular response.

Funding: The study was carried out within the framework of the state assignment of NIIKEL – Branch of FGBNU FIC ICiG SB RAS under the project No. FWNR-2024-0005.

Список литературы/References

1. Швецова А.М., Ершов К.И., Королев М.А., Мадонов П.Г., Трубина Л.В. Способ иммобилизации гиалуронидазы для получения пероральной лекарственной формы. *Сибирский научный медицинский журнал*. 2024;44(2):73-79. doi 10.18699/SSMJ20240209

[Shvetsova A.M., Ershov K.I., Korolev M.A., Madonov P.G., Trubina L.V. Method of hyaluronidase immobilization to produce an oral dosage form. *Sibirskij nauchnyj medicinskij zhurnal = Siberian Scientific Medical Journal*. 2024;44(2):73-79. doi 10.18699/SSMJ2024020 (in Russian)]

2. Куценко С.А. Основы токсикологии. СПб.: Наука, 2002
[Kutsenko S.A. Fundamentals of toxicology. SPb.: Nauka, 2002 (in Russian)]
3. Данченко Е.О. Оценка цитотоксичности фармацевтических субстанций с использованием клеточных культур. *Иммунология, аллергология, инфектология*. 2012;2:22-31
[Danchenko E.O. Evaluation of cytotoxicity of pharmaceutical substances using cell cultures. *Immunopathology, Allergology, Infectology*. 2012;2:22-31 (in Russian)]
4. Мазеркина И.А. Оценка лекарственной гепатотоксичности *in vitro* на клеточных моделях (обзор). *Безопасность и риск фармакотерапии*. 2023;11(2):131-144. doi 10.30895/2312-7821-2023-11-2-351
[Mazerkina I.A. In vitro assessment of drug-induced liver injury using cell-based models: a review. *Safety and Risk of Pharmacotherapy*. 2023;11(2):131-144. doi 10.30895/2312-7821-2023-11-2-351 (in Russian)]
5. Onakpoya I.J., Heneghan C.J., Aronson J.K. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. *BMC Med*. 2016;4(14):10. doi 10.1186/s12916-016-0553-2
6. Голубинская Е.П., Филоненко Т.Г., Ермола Ю.А. и др. Ультраструктурные особенности компонентов аэрогематического барьера легких при фиброзно-кавернозном туберкулезе легких. *Медицинский вестник Северного Кавказа*. 2019;14(1.2):180-185. doi 10.14300/mnnc.2019.14010
[Golubinskaya E.P., Filonenko T.G., Ermola Yu.A. et al. Ultrastructural features of the components of aerogematic lung barrier at fibro-cavernous tuberculosis. *Medical News of North Caucasus*. 2019;14(1.2):180-185. doi 10.14300/mnnc.2019.14010 (in Russian)]
7. Постникова О.А. Структурный анализ взаимодействия гепатоцитов, эндотелиоцитов и звездчатых клеток печени при вибрационном и вирусном взаимодействиях. Автореф. дис. докт. биол. наук. Новосибирск, 2012
[Postnikova O.A. Structural analysis of interaction of hepatocytes, endotheliocytes and stellate cells of the liver at vibration and viral interactions. Abstract of thesis … doct. biol. sci. Novosibirsk, 2012 (in Russian)]
8. Пупышев А.Б. Репаративная аутофагия и аутофаговая гибель клетки. Функциональные и регуляторные аспекты. *Цитология*. 2014;56(3):179-196
[Pupyshev A.B. Reparative autophagy and autophagic cell death. Functional and regulatory aspects. *Cytology*. 2014;56(3):179-196 (in Russian)]

doi 10.18699/bgrs2024-9.4-13

Внеклеточные везикулы глиальных клеток-предшественников человека обладают нейропротекцией путем модуляции пути PI3K-Akt при глутаматной эксайтотоксичности

Шеденкова М.О.[1, 2]*, Салихова Д.И.[1, 2], Некрасова А.А.[3], Бакаева З.В.[3], Гольдштейн Д.В.[1, 2]

[1] *НИИ молекулярной и клеточной медицины медицинского института ФГАОУ ВО «Российский университет дружбы народов», Москва, Россия*

[2] *ФГБНУ «Медико-генетический научный центр имени Н.П. Бочкова», Москва, Россия*

[3] *ФГАУ «НМИЦ здоровья детей» Минздрава России, Москва, Россия*

* *margarita.shedenkova@gmail.com*

Ключевые слова: внеклеточные везикулы; глиальные клетки-предшественники; протеом; нейропротекция; глутаматная эксайтотоксичность; иПСК

Мотивация и цель: Глутамат является одним из основных нейромедиаторов головного мозга. Его действие происходит через метаботропные или ионотропные глутаматные рецепторы, запуская разнообразные каскады, контролирующие процессы внутри клеток. Нарушение гомеостаза глутамата часто сопровождается развитием патологии нервной системы, приводящей к неврологическим и нейродегенеративным заболеваниям [1]. На сегодняшний день не существует препаратов, способных нивелировать токсическое действие глутамата, поэтому поиск новых нейропротективных средств остается актуальной задачей. Сравнительно недавно в группу нейротерапевтических агентов вошли внеклеточные везикулы, благодаря их способности к сохранению внутреннего содержимого и адресному связыванию с целевыми клетками. Целью данного исследования является изучение механизмов нейропротективного действия внеклеточных везикул, полученных из глиальных производных индуцированных плюрипотентных стволовых клеток человека (ВВ-ГКП), с помощью комбинированной работы протеомного анализа *in silico* и ингибиторного анализа *in vitro* на модели глутаматной эксайтотоксичности.

Методы и алгоритмы: Культуры глиальных клеток-предшественников были получены ранее от здорового донора путем дифференцировки индуцированных плюрипотентных стволовых клеток в глиальном направлении [2]. ВВ-ГКП получали и анализировали согласно методикам, описанным ранее [3]. Для постановки модели глутаматной эксайтотоксичности использовали первичную культуру кортикальных нейронов из крысят линии Wistar (P0-1). На 10-е сутки культивирования добавляли препараты ВВ-ГКП. На следующий день осуществляли постановку модели глутаматной эксайтотоксичности, согласно указанной методике [4]. Оценку выживаемости проводили с помощью МТТ-теста. Протеомный анализ ВВ-ГКП проводили с использованием хроматографической системы Ultimate 3000 Nano LC System, сопряженной с масс-спектрометром Q Exactive HF. Масс-спектрометрические данные сохранялись при автоматическом переключении между MS1 сканированием и вплоть до 15 MS/MS-сканирований

(метод topN). Масс-спектрометрические данные с прибора были конвертированы в масс-листы формата MGF (Mascot Generic Format) с помощью MSConvert (ProteoWizard Software Foundation). Для более тщательной идентификации белков был сгенерирован список пиков, который был проанализирован программами MASCOT (версия 2.5.1) и X! Tandem (ALANINE, 2017.02.01) с использованием базы данных UniProtKB, таксон Homo Sapiens. Проверку статистической достоверности идентификаций осуществляли на основании поиска по реверсированной базе данных белковых последовательностей (decoy reversed database). Полученный список белков анализировали с помощью программы String 8.11. Для подтверждения активации обнаруженного протеомным анализом пути выживания – PI3K-Akt использовали селективный ингибитор субъединицы PI3Kγ – AS605240 (1мкг/мл). Для статистической обработки данных МТТ-теста (односторонний тест ANOVA с пост-хок тестом Холм-Шидака и поправкой на множественные сравнения) использовали программу SigmaPlot 12.5. Данные представлены в виде среднего значения и стандартного отклонения.

Результаты: При анализе биоинформатических результатов белки были категоризированы согласно их локализации и участию в биологических процессах. По локализации 60 % идентифицированных белков относится к группе «внеклеточные везикулы». Это позволяет предположить, что был получен препарат, обогащенный ВВ-ГКП. 13 и 5 % обнаруженных белков, относятся к группам «белковый мембранный комплекс» и «везикулярное окаймление», которые возможно локализованы в мембранах самих везикул. Кроме того, 8, 7 и 4 % белков были отнесены к категориям «люмен секреторных гранул», «мембрана Гольджи» и «поздняя эндосома» соответственно (особенно разнообразные белки семейства Rab (из суперсемейства Ras малых ГТФазных белков). Среди групп биологических процессов наиболее представленными, являются категории: «метаболические процессы» (60 %), «процессы иммунной системы» (28 %) и «клеточная коммуникация» (26 %). Также, были выделены группы: «негативная регуляция клеточной гибели» (9 %), например, белок 14-3-3, являющийся прямым ингибитором белка Bad; «окислительно-восстановительные процессы» (8 %); «ответ на снижение уровня кислорода» (6 %), в которых были обнаружены разнообразные пероксиредоксины: пероксиредоксин-6, пероксиредоксин-4; «позитивная регуляция нейрогенеза» (5 %); «NIK-Nf-kB сигналинг» (4 %), с идентифицированными белками, например, субъединицы ядерного фактора «каппа-би» (Nf-kB) комплекса; «регенерация» (2 %).

Также был проведен анализ белков по различным базам данных: KEGG и Reactome pathways. Среди установленных сигнальных путей KEGG стоит особо отметить «сигнальный путь PI3K-Akt», который является одним из основных путей, способствующих выживанию нейроглиальных клеток при нейротоксичности. Из проанализированнх белков был выделен 51 белок, относящийся к данному пути, например: субъединицы рецепторов ITGA и ITGB, а также обе субъединицы HSP90 и все субъединицы белка Ras, которые состоят в комплексе активатора AKT. Кроме того, были обнаружены пути «MAPK сигналинг» (2,3 %), «mTOR сигналинг» (2 %), «HIF-1 сигналинг» (2 %), «аппелин сигналинг» (2 %), «релаксин сигналинг» (2 %), «AMPK сигналинг» (2 %), нейротрофиновый сигналинг (1 %) и «TGF-beta сигналинг» (1 %), которые также принимают участие в нейропротекции. На основании базы данных Reactome pathways были выделены следующие категории: «метаболизм» (24 %), «иммунная

систем» (23 %), «развитие нервной системы» (13 %), «навигация аксонов» (13 %), «гемостаз» (9 %), «сигналинг с помощью рецепторов ROBO» (около 9 %), «PIP3 активируемые сигналы AKT» (4 %), «рецепторы нейротрансмиттеров и постсинаптическая передача сигнала» (2 %), «сигналинг EPH-Ephrin» (2 %), «сигналинг под действием VEGF» (2 %), «сигнализация с помощью NTRKs» (1.5 %), «активация MAP2K и MAPK» (1 %), «детоксикация реактивных форм кислорода» (1 %) и «mTORC1-опосредованная сигнализация» (1 %).

На модели глутаматной эксайтотоксичности наблюдали снижение числа жизнеспособных кортикальных нейронов на 37.5 ± 4.03 % при добавлении 100 мкМ глутамата по сравнению с группой контроля, которая принималась за 100 %. Добавление препарата ВВ-ГКП (10 мкг/мл) достоверно увеличивало количество жизнеспособных клеток до контрольных значений ($p < 0.001$). Также было установлено, что присутствие AS605240 (1мкг/мл) одновременно с ВВ-ГКП полностью нивелировало нейропротекторное действие препарата ($p < 0.001$). Это позволяет предположить, что основным из активируемых путей выживания нейронов является именно PI3K-Akt сигнальный путь.

Выводы: Полученные результаты свидетельствуют о том, что ВВ-ГКП обладают нейропротекторным действием при глутаматной эксайтотоксичности благодаря содержащимся в них белкам, снижающих гибель нейронов с помощью различных механизмов и сигнальных путей, среди которых подтверждено участие сигнального пути PI3K-Akt. Эти данные позволяют предположить, что ВВ-ГКП могут стать новым видом терапевтического препарата для лечения неврологических заболеваний.

Финансирование: Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (проект № КБК 075 0110 47 1 S7 24600 621) по теме «Разработка новых лекарственных средств для терапии неврологических заболеваний».

Extracellular vesicles of human glial progenitor cells exhibit neuroprotection by modulating the PI3K-Akt pathway in glutamate excitotoxicity

Shedenkova M.O.[1, 2]*, Salikhova D.I.[1, 2], Nekrasova A.A.[3], Bakaeva Z.V.[3], Goldshtein D.V.[1, 2]

[1] *RUDN University, Moscow, Russia*

[2] *Research Centre for Medical Genetics, Moscow, Russia*

[3] *National Medical Research Center for Children's Health, Moscow, Russia*

* *margarita.shedenkova@gmail.com*

Key words: extracellular vesicles; glial progenitor cells; proteome; neuroprotection; glutamate excitotoxicity; iPSCs

Motivation and Aim: Glutamate is one of the main neurotransmitters in the brain. Its action occurs through metabotropic or ionotropic glutamate receptors, triggering a variety of cascades that control processes inside cells. Disruption of glutamate homeostasis is often accompanied by the development of nervous system pathology leading to neurological and neurodegenerative diseases [1]. To date, there are no drugs

capable of leveling the toxic effect of glutamate, so the search for new neuroprotective agents remains an urgent task. Comparatively recently, extracellular vesicles have joined the group of neurotherapeutic agents because of their ability to preserve their internal content and targeted binding to target cells. The aim of this study is to investigate the mechanisms of neuroprotective action of extracellular vesicles derived from glial progenitor cells (EV-GPC) of human induced pluripotent stem cells using the combined work of *in silico* proteomic analysis and *in vitro* inhibitor assay in a model of glutamate excitotoxicity.

Methods and Algorithms: Glial progenitor cell cultures were obtained previously from a healthy donor by differentiation of induced pluripotent stem cells in the glial direction [2]. EV-GPCs were obtained and analyzed according to the methods described previously [3]. Primary culture of cortical neurons from Wistar rats (P0-1) was used to set up a model of glutamate excitotoxicity. On the 10th day of cultivation EV-GPC drugs were added. On the next day, the model of glutamate excitotoxicity was performed according to the indicated method [4]. Survival was assessed using the MTT test.

Proteomic analysis of EV-GPC was performed using an Ultimate 3000 Nano LC System chromatography system interfaced with a Q Exactive HF mass spectrometer. Mass spectrometric data were stored by automatically switching between MS1 scans and up to 15 MS/MS scans (topN method). Mass spectrometric data from the instrument were converted to Mascot Generic Format (MGF) mass sheets using MSConvert (ProteoWizard Software Foundation). For more thorough protein identification, a peak list was generated and analyzed by MASCOT (version 2.5.1) and X! Tandem (ALANINE, 2017.02.01) using the UniProtKB database, taxon Homo Sapiens. The statistical reliability of the identifications was checked by searching the decoy reversed database (decoy reversed database). The obtained list of proteins was analyzed using String 8.11 program. To confirm the activation of the PI3K-Akt survival pathway detected by proteomic analysis, a selective inhibitor of PI3Kγ subunit – AS605240 (1μg/ml) was used. SigmaPlot 12.5 program was used for statistical processing of MTT-test data (one-way ANOVA with Holm-Shidak post-hoc test and correction for multiple comparisons). Data are presented as mean value and standard deviation.

Results: In the analysis of bioinformatic results, proteins were categorized according to their localization and involvement in biological processes. Based on localization, 60 % of the identified proteins belonged to the “extracellular vesicles” group. This suggests that a EV-GPC-enriched drug was obtained. 13 and 5 % of the detected proteins, belong to the groups “protein membrane complex” and “vesicular fringing”, which are possibly localized in the membranes of the vesicles themselves. In addition, 8, 7, and 4 % of the proteins were categorized as “secretory granule lumen”, “Golgi membrane”, and “late endosome”, respectively (especially the diverse Rab family proteins (from the Ras superfamily of small GTPase proteins). Among the groups of biological processes, the most represented categories are: “metabolic processes” (60 %), “immune system processes” (28 %) and “cell communication” (26 %). Also, next groups were identified: “negative regulation of cell death” (9 %), such as protein 14-3-3, which is a direct inhibitor of Bad protein; “redox processes” (8 %); “response to reduced oxygen levels” (6 %), in which a variety of peroxiredoxins were found: peroxiredoxin-6, peroxiredoxin-4; “positive regulation of neurogenesis” (5 %); “NIK-Nf-kB signaling” (4 %), with identified proteins such as the subunit of the nuclear factor kappa-bi (Nf-kB) complex; “regeneration” (2 %).

Proteins were also analyzed against different databases: KEGG and Reactome pathways. Among the identified KEGG signaling pathways, the “PI3K-Akt signaling pathway” is of particular note, which is one of the main pathways contributing to the survival of neuroglial cells during neurotoxicity. From the proteins analyzed, 51 proteins belong to this pathway, such as: ITGA and ITGB receptor subunits, as well as both subunits of HSP90 and all subunits of Ras protein, which are members of the AKT activator complex. In addition, we found MARK signaling (2.3 %), mTOR signaling (2 %), HIF-1 signaling (2 %), appelin signaling (2 %), relaxin signaling (2 %), AMRK signaling (2 %), neurotrophin signaling (1 %), and TGF-beta signaling (1 %) pathways that are also involved in neuroprotection. Based on the Reactome pathways database, the following categories were identified: “metabolism” (24 %), “immune system” (23 %), “nervous system development” (13 %), “axon navigation” (13 %), “hemostasis” (9 %), “ROBO receptor-mediated signaling” (about 9 %), “PIP3-activated AKT signaling” (4 %), “neurotransmitter receptors and postsynaptic signal transduction” (2 %), “EPH-Ephrin signaling” (2 %), “VEGF-activated signaling” (2 %), “NTRKs-mediated signaling” (1.5 %), “MAP2K and MAPK activation” (1 %), “detoxification of reactive oxygen species” (1 %), and “mTORC1-mediated signaling” (1 %).

In the glutamate excitotoxicity model, a 37.5±4.03 % decrease in the number of viable cortical neurons was observed when 100 μM glutamate was added compared to the control group (100 %). The addition of EV-GPC (10 μg/mL) significantly increased the number of viable cells to control values ($p<0.001$). It was also found that the presence of AS605240 (1μg/mL) simultaneously with EV-GPC completely negated the neuroprotective effect of the drug ($p<0.001$). This suggests that it is the PI3K-Akt signaling pathway that is the major neuronal survival pathway activated.

Conclusions: The results obtained indicate that EV-GPC have a neuroprotective effect in glutamate excitotoxicity because of the proteins they contain that reduce neuronal death through various mechanisms and signaling pathways, among which the involvement of the PI3K-Akt signaling pathway was confirmed. These data suggest that EV-GPC may become a new type of therapeutic drug to treat neurological diseases.

Funding: The study is supported by the Ministry of Education and Science of the Russian Federation (project No. KBK 075 0110 47 1 S7 24600 621) on the topic “Development of new drugs for the therapy of neurological diseases”.

Список литературы/References

1. Andersen J.V. et al. Glutamate metabolism and recycling at the excitatory synapse in health and neurodegeneration. *Neuropharmacology.* 2021;196:108719
2. Salikhova D. et al. Therapeutic effects of hipsc-derived glial and neuronal progenitor cells-conditioned medium in experimental ischemic stroke in rats. *Int J Mol Sci.* 2021;22(9):4694
3. Turovsky E.A. et al. Mesenchymal stromal cell-derived extracellular vesicles afford neuroprotection by modulating PI3K/AKT pathway and calcium oscillations. *Int J Biol Sci.* 2022;18(14):5345-5368
   Zgodova A. et al. Isoliquiritigenin Protects Neuronal Cells against Glutamate Excitotoxicity. *Membranes*. 2022;12(11):1052

9 Симпозиум «Биомедицина, биоинформатика и системная компьютерная биология»

Symposium “Biomedicine, bioinformatics and systems computational biology”

doi 10.18699/bgrs2024-9.5-01

Планирование эксперимента в области трехмерной печати биосовместимых скаффолдов с помощью модели активного машинного обучения

Борковская Е.В.[1]*, Вилински-Мазур К.А.[2], Коломенский Д.С.[2], Кириллов Б.А.[2, 3]

[1] *Московский физико-технологический институт, Долгопрудный, Россия*

[2] *Центр технологий материалов, Сколковский институт науки и технологий, Москва, Россия*

[3] *Центр высокоточного редактирования и генетических технологий для биомедицины, Институт биологии гена Российской академии наук, Москва, Россия*

** borkovskaia.ev@phystech.edu*

Ключевые слова: активное обучение; машинное обучение; 3D-биопечать; биосовместимые скаффолды

Мотивация и цель: Технологии трехмерной биопечати стремительно развиваются, предоставляя новые возможности для регенеративной медицины и тканевой инженерии [1]. Кроме того, трехмерные биомодели, созданные в лабораторных условиях, становятся важным инструментом для изучения механизмов заболеваний и тестирования новых лекарственных средств в условиях, максимально приближенных к естественным [2]. Основной проблемой при создании биосовместимых скаффолдов, композитных пористых структур из гидрогеля, заселяемых клетками, остается оптимизация параметров печати. Применение методов активного машинного обучения может значительно сократить число экспериментальных итераций для достижения необходимых качества и эффективности печати.

Методы и алгоритмы: Для оптимизации параметров экструзионной 3D-печати биосовместимых скаффолдов разработана модель активного машинного обучения [3]. Модель выбирает наиболее информативные эксперименты для улучшения качества печати на основе своей собственной оценки уверенности предсказания, после чего параметры печати, такие как температура шприца, скорость печати и состав гидрогеля, варьируются в соответствии с рекомендациями модели.

Результаты: Применение разработанной модели активного машинного обучения позволило улучшить качество биосовместимых скаффолдов, сократив при этом количество необходимых экспериментальных итераций.

Выводы: Использование активного машинного обучения для планирования экспериментов в 3D-биопечати позволяет существенно повысить эффективность процесса и качество получаемых скаффолдов. Предлагаемый подход способствует более быстрому и точному подбору параметров печати. Результаты исследования могут быть использованы для разработки и внедрения индивидуализированных решений в биомедицинских исследованиях и клинической практике.

Active machine learning for experimental planning for 3D printing of biocompatible scaffolds

Borkovskaya E.[1]*, Vilinski-Mazur K.[2], Kolomenskiy D.[2], Kirillov B.[2, 3]

[1] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

[2] *Center of Material Technologies, Skolkovo Institute of Science and Technology, Moscow, Russia*

[3] *Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia*

** borkovskaia.ev@phystech.edu*

Key words: active learning; machine learning; 3D bioprinting; biocompatible scaffolds

Motivation and aim: Technologies of 3D bioprinting currently in rapid development provide new opportunities for regenerative medicine and tissue engineering [1]. Moreover, lab-made three-dimensional biomodels have become an important tool for studying disease dynamics and testing of new remedies in environments that closely resemble natural ones [2]. However, the need for search of optimal printing parameters remains main challenge in creation of biocompatible scaffolds – composite porous structures made of hydrogel and infused with cells. Application of active machine learning methods [3] allows to reduce the number of experimental iterations needed for achieving required quality and effectiveness of bioprinting.
Methods and algorithms: We develop a model based on active machine learning [3] for optimization of parameters for printing biocompatible scaffolds with extrusion 3D bioprinter. The model selects the most informative experiments for improvement of printing quality based on its internal estimation of prediction confidence. After that we change the printing parameters, such as syringe temperature, printing speed and hydrogel composition during the experiment based on the recommendations given by the model.
Results: Application of the active learning-based model allows to improve the quality of biocompatible scaffolds while reducing the number of required experimental iterations.
Conclusions: Application of active machine learning for experimental planning in 3D bioprinting allows to significantly improve the effectiveness of printing and quality of the resulting scaffolds. This approach speeds up parameter selection for accurate printing. The results of the current research may be applied towards development of individualized biomedical solutions and their translation into clinical practice.

Список литературы/References

1. Ding Z., Tang N., Huang J., Cao X., Wu S. Global hotspots and emerging trends in 3D bioprinting research. *Front Bioeng Biotechnol*. 2023;11:1169893. doi 10.3389/fbioe.2023.1169893
2. Gao G., Ahn M., Cho W.W., Kim B.S., Cho D.W. 3D Printing of pharmaceutical application: drug screening and drug delivery. *Pharmaceutics.* 2021;13(9):1373. doi 10.3390/pharmaceutics13091373
3. Wan T. et al. A survey of deep active learning for foundation models. *Intell Comput*. 2023;2:0058. doi 10.34133/icomputing.0058

doi 10.18699/bgrs2024-9.5-02

Материалы и технологии изготовления перспективных сердечно-сосудистых имплантов

Лактионов П.П.[1, 2]*, Степанова А.О.[1, 2], Назаркина Ж.К.[1], Челобанов Б.П.[1], Кузьмин И.Е.[1], Мурашев И.С.[2], Осипова О.С.[2], Гостев А.А.[2], Карпенко А.А.[2], Черноносова В.С.[1, 2]

[1] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

[2] *Национальный медицинский исследовательский центр имени академика Е.Н. Мешалкина, Новосибирск, Россия*

* *lakt@niboch.nsc.ru*

Ключевые слова: атеросклероз; реканализация артерий; материалы для сердечно-сосудистой хирургии; электроспиннинг; полиуретаны

Мотивация и цель: Распространенность, прогрессирующее течение атеросклероза и сопутствующего сахарного диабета в последние десятилетия приводят к росту окклюзионных поражений коронарных и периферических артерий малого диаметра (до 6 мм). Их распространенность в общей популяции составляет от 3 до 10 % и возрастает до 20 % у людей старше 70 лет [1]. Заболевания периферических артерий оказывают существенное влияние на качество жизни больных и при прогрессировании являются основной причиной их инвалидизации и смертности. В связи с этим разработка новых материалов для сердечно-сосудистой хирургии (ССХ), которые позволили бы повысить эффективность хирургических процедур является актуальной задачей.

Наилучшим материалом для протезирования сосудов являются аутовены, проходимость которых в кратковременной и длительной перспективе превосходит таковую у синтетических протезов сосудов (ПС) и у эндоваскулярных технологий реканализации артерий [2]. Однако пригодные для хирургии аутовенозные трансплантаты доступны не более, чем у 20 % больных, а децеллюляризованные ксеногенные ПС, синтетические ПС, изготовленные из экспандированного политетрафторэтилена (эПТФЭ) или полиэтилентерефталата недостаточно эффективны, поскольку склонны к формированию аневризм, обширной неоинтимы, и имеют высокую частоту рестенозов и тромбозов [3]. Для того чтобы обеспечить долговременную состоятельность ПС, материалы, используемые для изготовления ПС и иных изделий для ССХ должны обладать достаточной механической прочностью, близкой к аналогу эластичностью/податливостью, биостабильностью, то есть способностью длительное время сохранять прочность после имплантации (рис. 1).

Кроме того, ПС должны выдерживать долговременную периодическую нагрузку без деформации и остаточного удлинения (аневризмы). Подходы, которые используются для повышения гемосовместимости, в кратковременной и долговременной перспективе, а также биосовместимости, необходимым требованием которой является способность поддерживать формирование однослойного непрерывного и не активированного эндотелия, минимизировать воспаление, гиперплазию рубцовой ткани (неоинтима, парапротезная капсула), представлены на рис. 2.

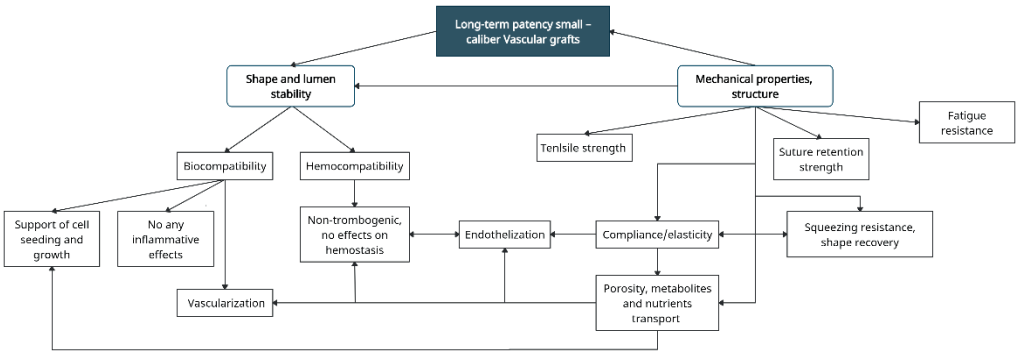


Рис. 1. Факторы, влияющие на долговременную состоятельность ПС

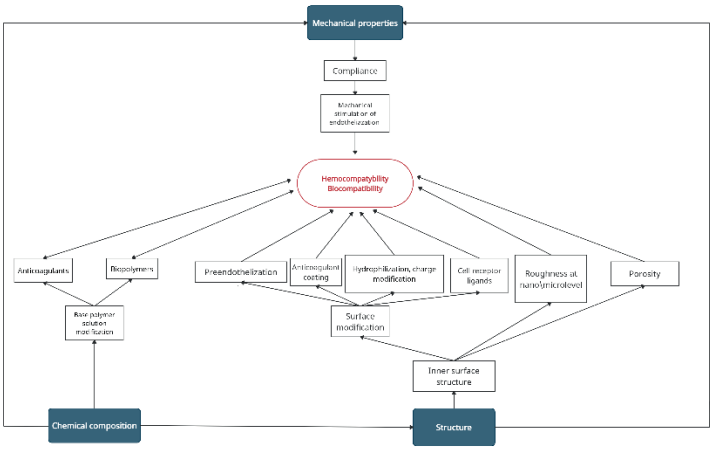


Рис. 2. Факторы и подходы повышения гемо- и биосовместимости ПС

Методы и алгоритмы: Для того чтобы изготовить такие изделия были использованы разные подходы, основанные как на использовании изделий из аллогенных (аутологичных) клеток, так и на изготовлении изделий из синтетических биодеградируемых, биостабильных (полигидроксиалконаты, полидиоксаноны, полигидроксисебокаты и т. д., более стабильных полимеров типа поликапролактона, еще более стабильных полиуретанов (ПУ), полисиликаты и гетерополимеры на их основе [4] или природных полимеров, их смесей, а также комбинаций клеточных технологий с заселяемыми полимерными конструкциями (рис. 3). Кроме технологии получения экспандированного ПТФЭ (эПТФЭ), который широко используется для изготовления изделий для ССХ, описано использование таких технологий как разделение фаз, вакуумная сушка, полимеризация гидрогелей, электроспиннинг (ЭС), разные варианты модификации поверхности изделий.

Мы исследовали ПС, изготовленные методом ЭС из белок-наполненных ПУ и стенты с лекарственно-наполненным покрытием, нанесенным методом ЭС в экспериментах *in vitro* и *in vivo*. Для покрытий стентов были использованы скэффолды на основе поликапролактона, с введенным в состав волокна паклитакселем или сиролимусом. ПС были изготовлены из полиэфирного (простой эфир) алифатического ПУ Tecoflex 80A, полиэфирного (простой эфир) ароматического ПУ Pellethane 2363-80A, полиэфирного (сложный эфир)

алифатического ПУ Tecothane TT-1074A, поликарбонатного алифатического ПУ Carbothane PC-3575A.

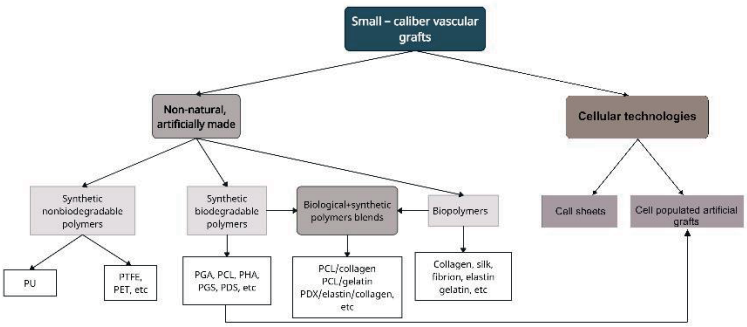


Рис. 3. Биологические, синтетические и комбинированные ПС

Результаты: На медицинском рынке уже представлены ПС, предназначенные для доступа при гемодиализе и изготовленные ЭС из полиуретанов от таких компаний как Nicast Ltd. (Израиль, протез AVflo™), Bard Inc. (США, протез VECTRA®) и Dialybrid S.r.l., (Италия, протез Silkothane®), а также изготовленные аналогичным образом стенты с покрытиями PK Papirus, (BIOTRONIK SE & Co. KG, Германия). Компания Humacyte Inc. (США) предложила изготавливать ПС на основе аутологичных клеток (изначально ГМК, потом ИПС) культивируемых на биодеградируемом матриксе в биореакторе [5]. В настоящее время такие ПС находятся на последней стадии клинических испытаний, хотя и не демонстрируют выдающихся показателей (проходимость 58 % через 5 лет).

Для изготовления лекарственно-наполненных покрытий стентов были подобраны составы растворов для ЭС, обеспечивающие хорошую гемо- и биосовместимость скэффолдов и условия упаковки лекарств в скэффолды с длительной, двухфазной доставкой препаратов. В экспериментах *in vivo* было показано, что стенты с лекарственным покрытием обладают очевидным преимуществом перед непокрытыми стентами [6, 7].

Был отработан состав внутреннего слоя ПС, который демонстрировал хорошую гемо- и биосовместимость (эти показатели не сильно отличались между разными типами ПУ). Показано, что наименее стабильный Tecoflex 80A в составе ПС практически не деградирует спустя 6 мес после имплантации в брюшную аорту крыс. Исследуемые ПС на основе ПУ значительно превосходят ПС из эПТФЭ по толщине неоинтимы, индукции воспаления и удобству установки. Исследовано влияние стерилизации на стабильность ПУ и подобраны оптимальные условия стерилизации изделий, а также разработано оборудование и технология для мелкосерийного изготовления ПС диаметром от 2 до 7 мм и длиной до 60 см [8–13].

Выводы: Литературные данные и результаты наших исследований свидетельствуют о высокой эффективности ПС и покрытий стентов, изготовленных ЭС из белок-наполненных полимеров. Полученные данные позволяют рекомендовать стенты с лекарственно-наполненным покрытием и ПС из белок-наполненного ПУ для доклинических и клинических испытаний.

Финансирование: Исследование поддержано финансированием ИХБФМ СО РАН по ГЗ № 121031300042-1 и частично из средств НМИЦ им. Е.Н. Мешалкина, ГЗ 121032300337-5.

Materials and technologies for manufacturing of the next generation cardiovascular implants

Laktionov P.P.[1, 2]*, Stepanova A.O.[1, 2], Nazarkina Zh.K.[1], Chelobanov B.P.[1], Kuzmin I.E.[1], Murashev I.S.[2], Osipova O.S.[2], Gostev A.A.[2], Karpenko A.A.[2], Chernonosova V.S.[1, 2]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *National Medical Research Center named after Academician E.N. Meshalkina, Novosibirsk, Russia*

* *lakt@niboch.nsc.ru*

Key words: atherosclerosis; recanalization of arteries; materials for cardiovascular surgery; electrospinning; polyurethanes

Motivation and Aim: The prevalence and progressive course of atherosclerosis and concomitant diabetes mellitus in recent decades have led to an increase in occlusive lesions of coronary and peripheral small diameter arteries (up to 6 mm). Their prevalence ranges from 3 to 10 %, and increases to 20 % in people over 70 years [1]. Peripheral arterial diseases have a significant impact on the quality of life and the main cause of patient's disability and mortality. In this regard, the development of new materials for cardiovascular surgery (CVS), which would improve the efficiency of surgical procedures, is an urgent task.

The best material for vascular prosthetics are patient's own veins, the patency of which in the short and long term is superior to that of synthetic vascular grafts (VG) and endovascular technologies for arterial recanalization [2]. However, autovenous grafts suitable for surgery are available in no more than 20 % of patients. Decellularized xenogeneic VG, synthetic VG made from expanded polytetrafluoroethylene (ePTFE) or polyethylene terephthalate are not effective enough, since they are prone to the formation of aneurysms, extensive neointima, and have a high incidence of restenosis and thrombosis [3]. In order to ensure the long-term patency of VG, the materials used for manufacture of VG and other devices for the CVS must have sufficient mechanical strength, elasticity/compliance close to the analogue, biostability, i.e. the ability to maintain strength for a long time after implantation (Fig. 1).

In addition, VG must withstand long-term pulsing load without deformation and residual elongation (aneurysm). Approaches that are used to increase hemocompatibility, in the short and long term, as well as biocompatibility, the necessary requirement of which is the ability to maintain the formation of a single-layer continuous and non-activated endothelium, minimize inflammation, hyperplasia of scar tissue (neointima, paraprosthetic capsule), are presented in Fig. 2.

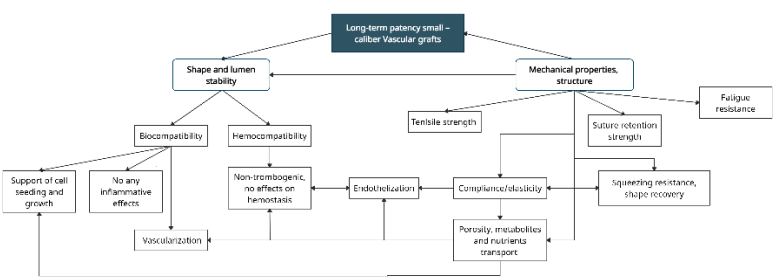


Fig. 1. Factors influencing the long-term viability of the VG

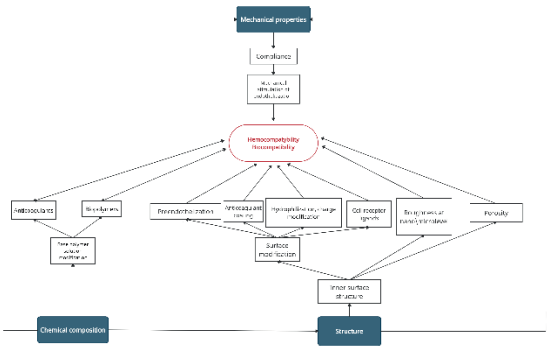


Fig. 2. Factors and approaches to increase hemo- and biocompatibility of VG

Methods and Algorithms: In order to produce such products, different approaches were used, based both on the use of products from allogeneic (autologous) cells, and on the manufacture of products from synthetic biodegradable, biostable (polyhydroxyalconates, polydioxanones, polyhydroxysebocates, etc., more stable polymers such as polycaprolactone, even more stable PU, polysilicates and heteropolymers based on them) [4] or natural polymers, their mixtures, as well as combinations of cellular technologies and synthetic scaffolds (Fig. 3). In addition to the technology for producing expanded PTFE (ePTFE), which is widely used for the manufacture of products for dry drying, the use of technologies such as phase separation, vacuum drying, polymerization of hydrogels, electrospinning (ES), and various options for modifying the surface of products is described.

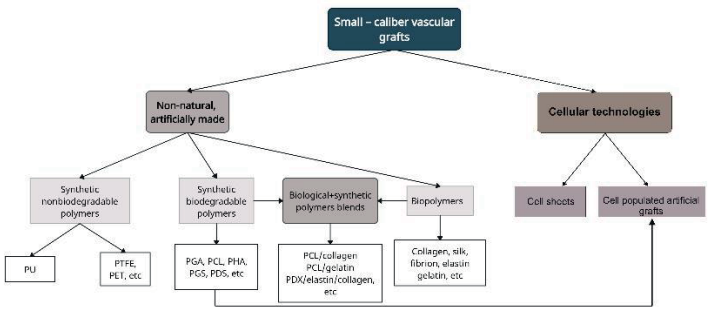


Fig. 3. Biological, synthetic and combined VG

We studied VGs manufactured by the ES method from protein-filled PUs and drug-loaded stents coated by the ES in vitro and *in vivo*. For stent coatings, scaffolds based on polycaprolactone were used, with paclitaxel or sirolimus incorporated into the fibers. The VGs were made from polyester (ether) aliphatic PU Tecoflex 80A, polyester (ether) aromatic PU Pellethane 2363-80A, polyester (ester) aliphatic PU Tecothane TT-1074A, polycarbonate aliphatic PU Carbothane PC-3575A.

Results: The medical market already offers VG intended for hemodialysis access and produced by ES from polyurethanes from Nicast Ltd. (Israel, AVflo™ prosthesis), Bard Inc. (USA, VECTRA® prosthesis) and Dialybrid S.r.l., (Italy, Silkothane® prosthesis), as well as similarly manufactured covered stents PK Papirus (BIOTRONIK SE & Co. KG, Germany). Humacyte Inc. (USA) proposed producing VG based on autologous cells

(initially SMCs, then iPS) cultured on a biodegradable scaffold in a bioreactor [5]. Currently, such VG are at the last stage of clinical trial, although they do not demonstrate outstanding performance (patency rate 58 % after 5 years).

To produce drug-loaded stent coatings, compositions of solutions for ES were selected to ensure good hemo- and biocompatibility of scaffolds and conditions for packaging drugs into scaffolds in order to ensure long-term, two-phase delivery of drugs. *In vivo* experiments have shown that drug-eluting stents have a clear advantage over bare stents [6, 7].

The composition of the inner layer of VGs was developed, which demonstrated good hemo- and biocompatibility (these indicators did not differ much between types of PU), it was shown that the least stable Tecoflex 80A practically does not degrade 6 months after VG implantation into the abdominal aorta of rats, all PU-based VG options are significantly superior to ePTFE VG in terms of neointimal thickness, inflammation induction, and ease of installation. The effect of sterilization on the stability of PU was studied and optimal conditions for VG sterilization were selected. Equipment and technology were developed for small-scale production of PU with a diameter of 2 to 7 mm and a length of up to 60 cm [8–13].

Conclusion: Published data and our own research data demonstrate the high efficiency of VG and stent coatings made by ES from protein-enriched polymers. The data obtained allow us to recommend drug-eluting stents and VG made from protein-enriched PU for preclinical and clinical trials.

Funding: The study was supported by funding from ICHBFM SB RAN 121031300042-1 and partially from the funds of the NM Research Center named after E.N. Meshalkin 121032300337-5.

Список литературы/References

1 Nehler M.R. et al. Epidemiology of peripheral arterial disease and critical limb ischemia in an insured national population. *J Vasc Surg*. 2014;60:686-695
2 Farber A. et.al. Surgery or endovascular therapy for chronic limb-threatening ischemia. *N Engl J Med.* 2022;387(25):2305-2316
3 Carrabba M., Madeddu P. Current strategies for the manufacture of small size tissue engineering vascular grafts. *Front Bioeng Biotechnol*. 2018;6:41. doi 10.3389/fbioe.2018.00041
4 Osipova O. et al. Potential of biodegradable synthetic polymers for use in small-diameter vessel engineering. *Macromol Res*. 2022;30:425-437
5 Kristofik N.J. et al. Improving *in vivo* outcomes of decellularized vascular grafts via incorporation of a novel extracellular matrix. *Biomaterials*. 2017;141:63-73
6 Kuznetsov K.A. et al. Vascular stents coated with electrospun drug-eluting material: functioning in rabbit iliac artery. *Polymers.* 2020;12(8):1741
7 Nazarkina Zh.K. et al. Sirolimus-eluting electrospun-produced matrices as coatings for vascular stents: dependence of drug release on matrix structure and composition of the external environment. *Materials.* 2020;13(12):2692
8 Chernonosova V.S. et al. Mechanical properties and biological behavior of 3D matrices produced by electrospinning from protein-enriched polyurethane. *Biomed Res Int.* 2018;2018:1380606
9 Chernonosova V.S. et al. Study of hemocompatibility and endothelial cell interaction of Tecoflex-based electrospun vascular grafts. *Intern J Polym Materials*. 2018;68(1-3):34-43. doi 10.1080/00914037.2018.1525721
10 Gostev A.A. et al. Electrospun polyurethane-based vascular grafts: physicochemical properties and functioning *in vivo*. *Biomed Mater*. 2019;15:015010
11 Gostev A.A. et al. *In vivo* stability of polyurethane-based electrospun vascular grafts in terms of chemistry and mechanics. *Polymers*. 2020;12:845
12 Chernonosova V.S. et al. Assessment of electrospun pellethane–based scaffolds for vascular tissue engineering. *Materials*. 2021;14(13):3678
13 Kuzmin I.E. et al. Method for manufacturing of small diameter vascular grafts. Russian patent № 2805590, date of priority 20.04.2023

doi 10.18699/bgrs2024-9.5-03

Разработка новых подходов к компартмент-локализованной долговременной доставке фармакологических препаратов

Назаркина Ж.К.[1]*, Челобанов Б.П.[1], Степанова А.О.[1], Карпенко А.А.[2], Лактионов П.П.[1, 2]

[1] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

[2] *Национальный медицинский исследовательский центр имени академика Е.Н. Мешалкина, Новосибирск, Россия*

** zha_naz@niboch.nsc.ru*

Ключевые слова: электроспиннинг; стенты с лекарственным покрытием; материалы для тканевой инженерии; сиролимус; паклитаксел; доставка лекарств; высвобождение лекарственных препаратов

Мотивация и цель: Поскольку сердечно-сосудистые заболевания (ССЗ) занимают первое место среди всех причин смертности и второе место среди причин инвалидизации населения РФ [1], актуальной задачей является разработка материалов, использование которых в сердечно-сосудистой хирургии позволило бы повысить эффективность хирургического вмешательства в кратковременной и долговременной перспективе.

Методы и алгоритмы: Электроспиннинг (ЭС) является уникальной технологией, позволяющей получать полимерные волокна из растворов полимеров, толщиной от десятков нанометров до нескольких микрон. В раствор для ЭС можно вводить не только полимеры, но и лекарственные препараты, наночастицы и функциональные биомолекулы. Варьируя состав раствора для ЭС можно контролировать высвобождение лекарственных препаратов из скэффолдов. Методом ЭС были получены лекарственно-наполненные скэффолды на основе поликапролактона (ПКЛ). В качестве лекарственных препаратов использовали сиролимус (СРЛ) и паклитаксел (ПТХ), обладающие антипролиферативным действием. Для векторизации доставки лекарственных препаратов были получены скаффолды на основе поликапролактона, содержащие волокна, наполненные наночастицами активированного угля (АУ). Для этого использовали уголь марки АХ-21 после последовательного измельчения на шаровой и бусинковой мельницах [2]. Структура полученных скэффолдов была исследована при помощи оптической микроскопии, сканирующей электронной микроскопии (СЭМ) и рентгеновской фотоэлектронной спектроскопии (РФЭС). Для исследования кинетики высвобождения препаратов были изготовлены скэффолды, содержащие радиоактивномеченые препараты. Изучено влияние деформации скэффолдов на высвобождение лекарств. Исследована био- и гемосовместимость полученных скэффолдов.

Результаты: Для изготовления лекарственно-наполненных скаффолдов в качестве базового полимера был использован поликапролактон, который характеризуется хорошими механическими характеристиками и биосовместимостью и нетоксичностью [3, 4]. Для улучшения гемосовместимости

в состав волокон добавляли человеческий сывороточный альбумин (ЧСА) [4]. Было изготовлено и исследовано три типа скэффолдов: ПКЛ, ПКЛ/ЧСА и ПКЛ/ЧСА/ДМСО, содержащих СРЛ или ПТХ. Наиболее длительное высвобождение лекарств наблюдалось для скэффолда ПКЛ/ЧСА/ДМСО. В течение 27 суток из данного типа скэффолдов высвобождалось не более 80% лекарственного препарата, что говорит о перспективности использования таких скэффолдов для пролонгированной доставки лекарств. Раскрытие стента при его имплантации приводит к растяжению покрытия стента, что может оказывать влияние на структуру скэффолда и характер высвобождения лекарств. Было исследовано влияние деформации скэффолдов на высвобождение лекарственных препаратов. Показано, что растяжение скэффолда в одном направлении приводит к ориентации волокон матрикса в направлении приложенной силы, без образования разрывов и без изменения диаметра волокон. Показано, что СРЛ и ПТХ быстрее высвобождаются в плазму крови (ПК) человека, чем в фосфатный буфер (PBS). При инкубации скэффолдов с ПК в динамических условиях скорость высвобождения выше. Для всех типов скэффолдов растяжение не оказывает значимого влияния на кинетику высвобождения лекарств из скэффолдов. Это свидетельствует о том, что деформация скэффолдов при установке стентов не влияет на скорость высвобождения препаратов.

Исследована кинетика высвобождения СРЛ и ПТХ из стентов, покрытых скаффолдами состава ПКЛ/ЧСА/ДМСО через артериальную стенку в PBS и в ПК. Стенты с покрытиями устанавливали на баллоны-катетеры (рис. 1), вводили их в свежеэксплатированную подвздошную артерию кролика, и исследовали высвобождение радиоактивного препарата через артериальную стенку. Полученные данные демонстрируют, что артериальная стенка эффективно удерживает лекарственные средства. ПТХ высвобождается через артериальную стенку в PBS быстрее, чем в плазму. СРЛ имеет схожую кинетику высвобождения в ПК и PBS. Не более 20 % ПТХ, и не более 10 % СРЛ потенциально способных высвободиться из матрикса, диффундирует через артерию в ПК в течение 24 ч, тогда как большая часть препарата накапливается в артериальной стенке.

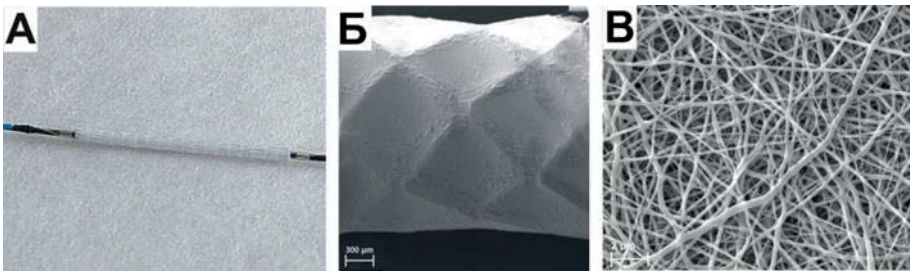


Рис. 1. Установленный на балон-катетер стент с покрытием ПКЛ/ЧСА/ДМСО/ПТХ, нанесенным методом электроспиннинга (А). Структура покрытия по данным СЭМ при увеличении 66× (Б), при увеличении 10000× (В)

Для векторизации доставки лекарственных препаратов в сосудистую стенку предлагается использовать скэффолды, содержащие несколько функциональных слоев: основной внешний (контактирующий со стенкой сосуда) слой, высвобождающий лекарственный препарат, и внутренний (ориентированный в

кровь) слой, содержащий наночастицы активированного угля и способный сорбировать лекарственный препарат. Сорбирующий слой необходим для уменьшения попадания лекарственного препарата в кровоток. Было исследовано 4 типа скэффолдов, изготовленных на основе ПКЛ, содержащих волокна (гомогенные и коаксиальные), наполненные наночастицами АУ. Показано, что кинетика связывания/высвобождения СРЛ из АУ-наполненных скэффолдов определяется составом волокон и отличается от кинетики, полученной при использовании свободного АУ. Показано, что клетки пупочной вены человека (HUVEC) и фибробласты десны человека (GF) адгезируют к скэффолдам на основе ПКЛ хуже, чем к культуральному пластику. Нагрузка волокон АУ не оказывает значительного влияния на адгезию и пролиферацию клеток, культивируемых на поверхности скэффолдов.

Для исследования гемосовместимости уголь-наполненных скэффолдов были выполнены тесты на гемолиз и адгезию тромбоцитов. Уровень гемолиза не превышал 1 %. Инкубация скэффолдов с обогащенной тромбоцитами плазмой показала, что адгезия тромбоцитов зависит от состава волокон. Минимальное количество прикрепившихся тромбоцитов наблюдалось для скэффолда с коаксиальными волокнами, в то время как для скэффолда, не содержащего частицы АУ, и скэффолда с гомогенными волокнами степень адгезии тромбоцитов была существенно выше. Таким образом, уголь-наполненный скэффолд с коаксиальными волокнами обладает повышенным атромбогенным потенциалом по сравнению с двумя другими скэффолдами.

Выводы: Предложены скэффолды на основе ПКЛ, в состав которых введены лекарственные средства. Исследована кинетика высвобождения лекарственных препаратов (паклитаксела и сиролимуса) из скэффолдов, в том числе при деформации. Независимо от состава волокон, растяжение скэффолдов не оказывает значимого влияния на кинетику высвобождения лекарств. Скэффолды на основе ПКЛ, содержащие ЧСА и ДМСО, являются наиболее перспективными в качестве покрытий для сосудистых стентов с длительным высвобождением лекарственных препаратов. Показано, что артериальная стенка эффективно удерживает лекарственные препараты. Накопление препарата в артериальной стенке способствует длительному поддержанию субцитотоксических концентраций лекарств в артериальной стенке и позволяют уменьшить дозу вводимых в покрытие стентов препаратов. Для решения проблемы вымывания лекарства током крови впервые было предложено использовать скэффолды, содержащие наполненный активированным углем барьерный слой. Показано, что АУ в составе барьерного слоя хорошо адсорбирует лекарственные средства, в том числе и в биологических жидкостях, при этом наполненные углем скаффолды не токсичны, хорошо совместимы с эукариотическими клетками и обладают хорошей гемосовместимостью. Скаффолды с уголь-наполненным слоем могут быть использованы для изготовления покрытий стентов, протезов сосудов, других имплантируемых трубчатых структур. Такое строение материала позволяет уменьшить расход лекарственных препаратов, уменьшить их системную доставку и повысить эффективность медицинских изделий.

Финансирование: Исследование поддержано бюджетным проектом по госзаданию (ГЗ 121031300042-1) и частично из средств НМИЦ им. акад. Е.Н. Мешалкина (ГЗ 121032300337-5).

New approaches of compartment-localized long-term drug delivery

Nazarkina Zh.[1]*, Chelobanov B.[1], Stepanova A.[1], Karpenko A.[2], Laktionov P.[1, 2]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *Meshalkin National Medical Research Center, Ministry of Health of the Russian Federation, Novosibirsk, Russia*

** zha_naz@niboch.nsc.ru*

Key words: electrospinning; drug-eluting stents; materials for tissue engineering; sirolimus; paclitaxel; drug delivery; controlled drug release

Motivation and Aim: Cardiovascular diseases (CVD) occupy the first place among all causes of death and the second place among the causes of disability of the population of the Russian Federation [1]. An urgent task is to develop materials for cardiovascular surgery that would improve the effectiveness of surgical intervention in the short and long term.

Methods and Algorithms: Electrospinning (ES) is a unique technology that allows the production of polymer fibers from polymer solutions ranging in thickness from tens of nanometers to several microns. Not only polymers can be introduced into the ES solution, but also drugs, nanoparticles and functional biomolecules. By varying the composition of the ES solution, the release of drugs from scaffolds can be controlled. Using the ES method, drug-filled scaffolds based on polycaprolactone (PCL) were obtained. Sirolimus (SRL) and paclitaxel (PTX), which have an antiproliferative effect, were used as drugs. To vectorize drug delivery, PCL-based scaffolds containing fibers filled with activated carbon (AC) nanoparticles were obtained. For this purpose, activated carbon AX-21 was used after sequential grinding with ball and bead mills [2]. The structure of the obtained scaffolds was studied using optical microscopy, scanning electron microscopy (SEM) and X-ray photoelectron spectroscopy (XPS). Scaffolds containing radioactively labeled drugs were manufactured to study the kinetics of drug release. The effect of scaffold deformation on drug release has been studied. The bio- and hemocompatibility of the obtained scaffolds has been investigated.

Results: For the manufacture of drug-filled scaffolds, polycaprolactone was used as the base polymer, which is characterized by good mechanical characteristics and biocompatibility and non-toxicity [3, 4]. To improve hemocompatibility, human serum albumin (HSA) was added to the fiber composition [4]. Three types of scaffolds were manufactured and investigated: PCL, PCL/HSA and PCL/HSA/DMSO containing SPL or PTX. The longest drug release was observed for PCL/HSA/DMSO scaffold. Within 27 days, no more than 80 % of the drug was released from this type of scaffolds. This indicates the prospects of using such scaffolds for prolonged drug delivery. The stent instalation leads to stretching of the stent coating, which can affect the scaffold structure and the drug release. The effect of scaffold deformation on drug release has been investigated. It is shown that the scaffold elongation leads to the orientation of the matrix fibers in the direction of the applied force, without the formation of breaks and without changing the diameter of the fibers. It has been shown that SRL and PTX are released faster into human blood plasma (BP) than into the phosphate buffer (PBS). When incubating scaffolds from a BP under dynamic conditions, the release rate is higher. For

all types of scaffolds, elongation has no significant effect on the drug release kinetics. This indicates that the deformation of scaffolds during the stent installation does not affect the rate of drug release.

The kinetics of the release of SRL and PTX from stents coated with PCL/HSA/DMSO scaffolds through the arterial wall in PBS and in BP have been studied. Coated stents were placed on balloon catheter (Fig. 1), injected into the freshly exposed iliac artery of a rabbit, and the release of a radioactive drug through the arterial wall was investigated. The data obtained demonstrate that the arterial wall effectively retains drugs. PTX is released through the arterial wall into PBS faster than into BP. SRL has similar release kinetics in BP and PBS. No more than 20% of the PTX, and no more than 10 % of the SRL released from the scaffolds, diffuses through the artery into BP within 24 hours, whereas most of the drug accumulates in the arterial wall.

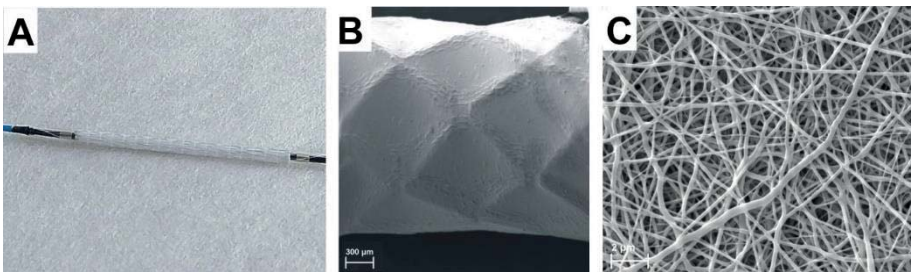


Fig. 1. An electrospun coated stent installed onto balloon catheter before (A). The structure of the coating according to the SEM data at an increase of 66 × (B), at an increase of 10000 × (C)

To vectorize drug delivery into the vascular wall, it is proposed to use scaffolds containing several functional layers: the main outer (in contact with the vessel wall) drug releasing layer, and the inner (blood-oriented) AC-containing layer capable of drug adsorbing. The adsorbing layer is necessary to reduce the drug penetration into the bloodstream. 4 types of PCL-based scaffolds containing homogeneous and coaxial AC-filled fibers were studied. It is shown that the kinetics of the binding/release of CRL from AC-filled scaffolds is determined by the fiber composition and differs from the kinetics obtained using free AC nanoparticles. It has been shown that human umbilical vein cells (HUVEC) and human gingival fibroblasts (GF) adhere to PCL-based scaffolds worse than to tissue culture plastic. The load of fibers with AC nanoparticles has no significant effect on the adhesion and proliferation of cells cultured on the surface of scaffolds.

To study the hemocompatibility of AC-filled scaffolds, tests for hemolysis and platelet adhesion were performed. The hemolysis level did not exceed 1%. Incubation of scaffolds with platelet-rich plasma showed that platelet adhesion depends on the fiber composition. The minimum number of attached platelets was observed for scaffold with coaxial fibers, while for scaffold without AC particles and scaffold with homogeneous fibers, the degree of platelet adhesion was significantly higher. Thus, a AC-filled scaffold with coaxial fibers has an increased athrombogenic potential compared to the other two scaffolds.

Conclusion: PCL-based scaffolds contain drugs (paclitaxel and sirolimus) have been proposed. Drug release kinetics from scaffolds, including during deformation, has been studied. Regardless of the fiber composition, scaffold elongation has no significant effect on the drug release kinetic. PCL-based scaffolds containing HSA and DMSO are the

most promising coatings for vascular stents with long-term drug release. It has been shown that the arterial wall effectively retains drugs. The drug accumulation in the arterial wall contributes to the long-term maintenance of subcitotoxic concentrations of drugs in the arterial wall and allows reducing the dose of drugs added into the stent coating. To solve the problem of drug washing out by blood flow, it was proposed to use scaffolds containing a barrier layer filled with activated carbon. It has been shown that AC in the barrier layer adsorbs drugs well, including in biological fluids, while carbon-filled scaffolds are non-toxic, well compatible with eukaryotic cells and have good hemocompatibility. Scaffolds with AC-filled layer can be used for the manufacture of stent coatings, vascular prostheses, and other implantable tubular structures. This structure of the material makes it possible to decrease the drug concentration, reduce their systemic delivery and increase the effectiveness of medical devices.

Funding: The study is supported Russian State-funded budget project of ICBFM SB RAS number 121031300042-1 and partially supported by the state assignment of the Ministry of Health of the Russian Federation number 121032300337-5.

Список литературы/References

1. Здравоохранение в России: Стат. сб. Росстат. М., 2023
[Healthcare in Russia: Stat. sat. Rosstat. Moscow, 2023 (in Russian)]
2. Nazarkina Z.K., Savostyanova T.A., Chelobanov B.P. et al. Activated carbon for drug delivery from composite biomaterials: the effect of grinding on sirolimus binding and release. *Pharmaceutics.* 2022;14(7):1386. doi 10.3390/pharmaceutics14071386
3. Mirbagheri M.S., Mohebbi-Kalhori D., Jirofti N. Evaluation of mechanical properties and medical applications of polycaprolactone small diameter artificial blood vessels. *Int J Basic Sci Med.* 2017;2:58-70
4. Chernonosova V.S., Kvon R.I., Stepanova A.O., Larichev Y.V., Karpenko A.A., Chelobanov B.P., Kiseleva E.V., Laktionov P.P. Human serum albumin in electrospun PCL fibers: structure, release, and exposure on fiber surface. *Polym Adv Technol.* 2017;28:819-827

doi 10.18699/bgrs2024-9.5-04

Введение нуклеиновых кислот в скэффолды: изготовление, исследование свойств и эффективности трансфекции культивируемых на них клеток

Черноносова В.[1]*, Хлебникова М.[1], Попова В.[1], Челобанов Б.[1], Киселева Е.[2], Байбородин С.[2], Дмитриенко Е.[1], Лактионов П.[1]

[1] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

* *vera_mal@niboch.nsc.ru*

Ключевые слова: электроспиннинг; скэффолды; нуклеиновые кислоты; наночастицы диоксида кремния; ДНК доставка; трансфекция клеток

Мотивация и цель: После имплантации тканеинженерные конструкции неизбежно контактируют с окружающими клетками. Для успешного функционирования такие импланты должны ингибировать реакцию на чужеродное тело (воспаление) и инициировать заселение конструкции клетками требуемого фенотипа, то есть управлять фенотипом близлежащих клеток. Для ингибирования воспаления может быть использовано введение в скэффолд низкомолекулярных лекарственных препаратов, в то время как модификация фенотипа требует локальной доставки в клетки не только низкомолекулярных соединений (например, вальпроевой кислоты), но и высокомолекулярных факторов роста, и/или нуклеиновых кислот кодирующих транскрипционные или ростовые факторы. Доставка в клетки кодирующих нуклеиновых кислот позволяет непосредственно влиять на фенотип трансфецируемых клеток и индуцировать экспрессию цитокинов и хемо аттрактантов и т. д. [1, 2].

В связи с этим задачей представленного исследования являлась разработка и исследование скэффолдов, нагруженных комплексами плазмидной ДНК (пДНК) с наночастицами диоксида кремния, которые предназначены для трансфекции ДНК в клетки, культивируемые на таких скэффолдах.

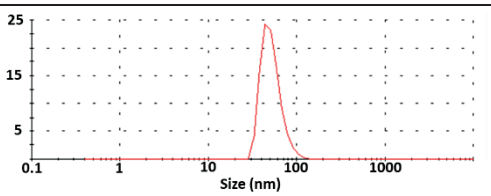
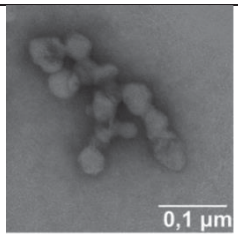
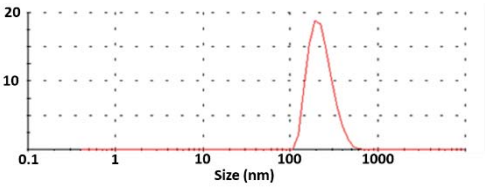
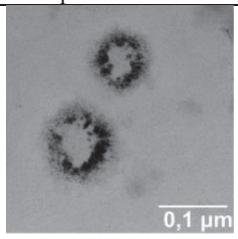
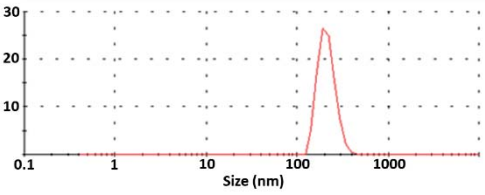
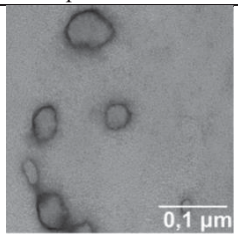
Методы и алгоритмы: Методом электроспиннинга были изготовлены двухслойные скэффолды. Основной слой скэффолда изготавливали из смеси 5 % поликапролактона (ПКЛ) (вес/объем) с 10 % желатина (жел) (вес/вес) в гексафторизопропаноле, а поверхностный слой изготавливали из аналогичной композиции с пДНК-наночастицами. ДНК-наночастицы представляли собой кремниевые наночастицы, на которые была сорбирована пДНК (КН-пДНК) или комплексы пДНК с полиэтиленимином в различной концентрации (КН-пДНК+ПЭИ). В качестве пДНК была использована phMGFP, свободная от ЛПС (менее 0.0125 Е/мг по данным LAL-теста). Размер пДНК-наночастиц был охарактеризован с помощью трансмиссионной электронной микроскопии (JEOL1400, Япония) и метода динамического рассеяния света (ДСР) (Zetasizer nano, Великобритания).

Физико-химические свойства скэффодлов были исследованы при помощи сканирующей электронной микроскопии (EVO 10, Германия) и разрывной машины Zwick/Roell Z10. Жизнеспособность клеток линии HEK293T определяли,

культивируя клетки на поверхности исследуемых скэффолдов в течение 2 и 5 суток с последующей оценкой количества живых клеток при помощи Alamar blue теста (Termo Fisher, США). Эффективность трансфекции клеток HEK293T плазмидой phMGFP оценивали с помощью флуоресцентной микроскопии (LSM 780 NLO, Германия) и количественной ПЦР (Bio-Rad, США).

Результаты: Для введения пДНК в скэффолды были использованы кремневые частицы, полученные по ранее описанной методике [3]. В предварительных экспериментах было обнаружено, что сорбционная емкость 1 мг КН составляет 1.4–1.5 мкг пДНК. Средний размер КН-пДНК полученный при помощи различных методов представлен в Таблице 1. Увеличение размеров КН-пДНК относительно исходных КН, свидетельствует о сорбции ДНК на их поверхности, а уменьшение их размеров после добавления ПЭИ указывает на конденсацию ДНК на поверхности КН.

Таблица 1. Характеризация КН различного состава с помощью ДСР и ТЭМ

	ДСР	ТЭМ
Исходные КН	 Диаметр = 52 ± 5 нм, ζ-потенциал = −2.26 ± 0.02 мВ	 Размер = 56.3 ± 14.6 нм
КН-пДНК	 Диаметр = 252 ± 45 нм, ζ-потенциал = −10.7 ± 0.2 мВ	 Размер= 77.9± 17.1 нм
КН-пДНК+ПЭИ9	 Диаметр = 205 ± 73 нм, ζ-потенциал = −4.55 ± 0.3 мВ	 Размер = 63.0± 9.7 нм

По данным СЭМ полученные скэффолды представляли собой волокнистые материалы с диаметром волокон 192 ÷ 356 нм и размером пор 5.0–7.2 мкм. Скэффолды с КН имели более шероховатую поверхность и наименьший диаметр волокон, по сравнению с контрольным скэффолдом ПКЛ-жел. Данные

механического теста указывают, что введение НК в состав полимерных волокон приводит к снижению их максимального удлинения до разрыва в 2 раза с 280 ± 32 до 132 ± 27 %, но не оказывает влияния на прочность скэффолдов, которая варьировалась от 4.18 ± 0.3 до 4.7 ± 0.48 МПа.

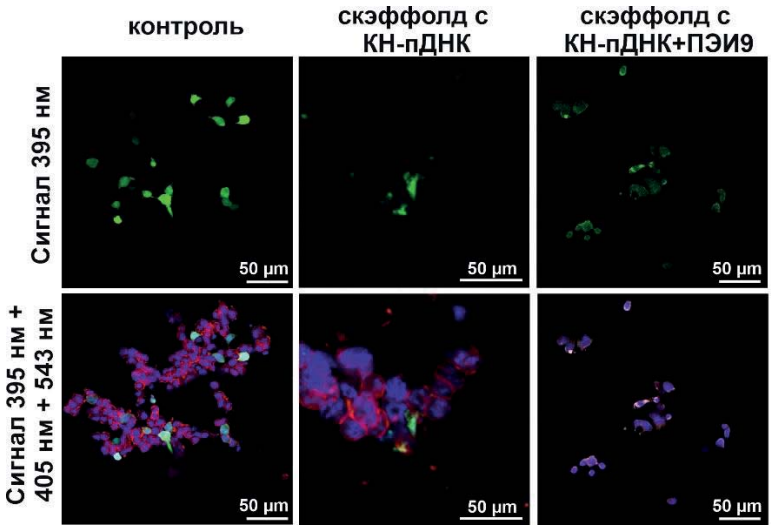


Рис. 1. Данные флуоресцентной микроскопии клеток HEK293T культивируемых на различных скэффолдах в течение 5 дней. Положительный контроль – клетки после трансфекции пДНК с использованием реагента TurboFect. Клетки окрашены Hoechst 33342 и Phalloidin-TRITC

пДНК-содержащие скэффолды не цитотоксичны в отношении культивируемых на них клеток линии HEK293T, но влияют на скорость пролиферации клеток. По данным флуоресцентной микроскопии достоверная трансфекция клеток HEK293T наблюдается при их культивировании на скэффолдах, содержащих КН-пДНК+ПЭИ, в отличие от скэффолдов, содержащих КН-пДНК (рис. 1). Данные флуоресцентной микроскопии коррелируют с данными количественной ПЦР экспрессии GFP мРНК (в качестве референтных генов использованы гены домашнего хозяйства 18S и GAPDH).

Выводы: Результаты работы свидетельствуют о возможности получения полимерных скэффолдов, изготовленных методом электроспиннинга и доставляющих нуклеиновые кислоты в культивируемые на них клетки с целью модификации их фенотипа/прямого репрограммирования.

Финансирование: Исследование поддержано государственным бюджетным проектом ИХБФМ СО РАН № 121031300042-1.

Introduction of nucleic acids into scaffolds: manufacturing, study of the properties and efficiency of cell transfection

Chernonosova V.[1]*, Khlebnikova M.[1], Popova V.[1], Chelobanov B.[1], Kiseleva E.[2], Baiborodin S.[2], Dmitrienko E.[1], Laktionov P.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *vera_mal@niboch.nsc.ru*

Key words: electrospinning; scaffolds; nucleic acids; nanoparticles of silicon dioxide; DNA delivery; cell transfection

Motivation and Aim: After implantation, tissue-engineered devices inevitably come into contact with surrounding cells. To function successfully, such implants must inhibit the reaction to a foreign body (inflammation) and initiate the population of the structure with cells of the required phenotype, i.e. control the phenotype of nearby cells. Low molecular weight drugs can be used to inhibit inflammation, while modification of the phenotype requires local delivery into cells of not only low molecular weight compounds (for example, valproic acid), but also high molecular weight growth factors, and/or nucleic acids encoding transcription or growth factors. Delivery of encoding nucleic acids into cells allows one to directly influence the phenotype of transfected cells and induce the expression of cytokines and chemoattractants, etc. [1, 2].

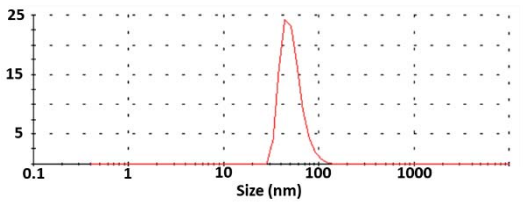
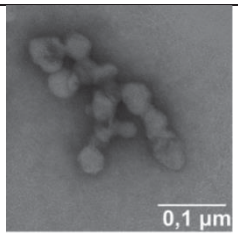
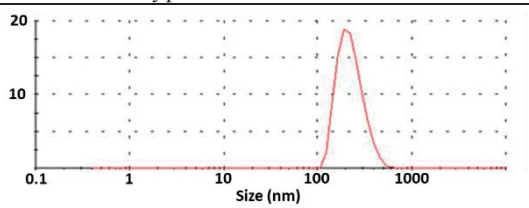
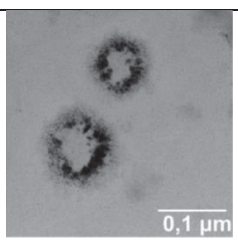
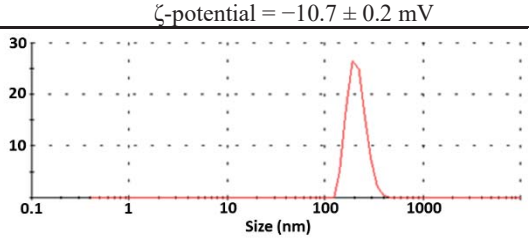
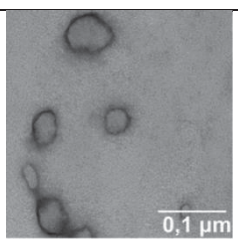
In this regard, the goal of the presented work was to develop and study scaffolds containing silica nanoparticles-associated plasmid DNA (pDNA) and intended for DNA transfection into cells cultured on such materials.

Methods and Algorithms: Two-layer scaffolds were fabricated using the electrospinning. The main layer of scaffold was made from a mixture of 5 % polycaprolactone (PCL) (w/v) with 10 % gelatin (gel) (w/w) in hexafluoroisopropanol, and the surface layer was made from a similar composition with different silica nanoparticles-associated pDNA (silica nanoparticles with pDNA (SiP-pDNA) or complexes of pDNA with polyethylenimine in various concentrations (SiP-pDNA + PEI)). phMGFP free of LPS was used as pDNA. The size of studied nanoparticles was characterized using transmission electron microscopy (JEOL1400, Japan) and dynamic light scattering (Zetasizer nano, UK).

The scaffolds were characterized using scanning electron microscopy (EVO 10, Germany) and stress-loading tests by machine Zwick/Roell Z10. The viability of HEK293T cells was determined by contact cells with the surface of the scaffolds for 2 and 5 days, followed by assessing the number of living cells using the Alamar blue test (Termo Fisher, USA). The efficiency of transfection of HEK293T cells with the phMGFP plasmid was assessed using fluorescence microscopy (LSM 780 NLO, Germany) and quantitative PCR (Bio-Rad, USA).

Results: The silica nanoparticles were prepared as described in [3]. In preliminary experiments, it was found that the sorption capacity of 1 mg of SiP is 1.4–1.5 μg of pDNA. The average size of SiP-pDNA obtained using various methods is presented in Table 1. An increase in the size of SiP-pDNA relative to the original SiP indicates the sorption of DNA on their surface, and a decrease in their size after the addition of PEI indicates DNA condensation on the surface of the nanoparticles.

Table 1. Characterization of different SiP using DLS and TEM

	Dynamic light scattering	TEM
SiP	 Diameter = 52 ± 5 nm ζ-potential = −2.26 ± 0.02 mV	 Size = 56.3 ± 14.6 nm
SiP-pDNA	 Diameter = 252 ± 45 nm ζ-potential = −10.7 ± 0.2 mV	 Size = 77.9± 17.1 nm
SiP-(pDNA+PEI9)	 Diameter = 205 ± 73 nm ζ-potential = −4.55 ± 0.3 mV	 Size = 63.0± 9.7 nm

According to SEM data, the scaffolds were fibrous materials with a fiber diameter of 192 ÷ 356 nm and a pore size of 5.0–7.2 µm. Scaffolds contained SiP had a rougher surface and a smaller fiber diameter compared to the control PCL-gel scaffold. Mechanical test data indicate that the introduction of SiP into the polymer composition leads to a decrease in their maximum elongation to break by 2 times from 280 ± 32 to 132 ± 27 %, but does not affect the strength of the scaffolds, which varied from 4.18 ± 0.3 to 4.7 ± 0.48 MPa.

pDNA-containing scaffolds are not cytotoxic to HEK293T cells cultured on them, but they do affect the rate of cell proliferation. According to fluorescence microscopy, reliable transfection of cells is observed when they are cultured on scaffolds containing SiP-(pDNA + PEI), in contrast to scaffolds containing SiP-pDNA (Fig. 1). Fluorescence microscopy data correlate with quantitative PCR data on GFP mRNA expression (housekeeping genes 18S and GAPDH were used as reference genes).

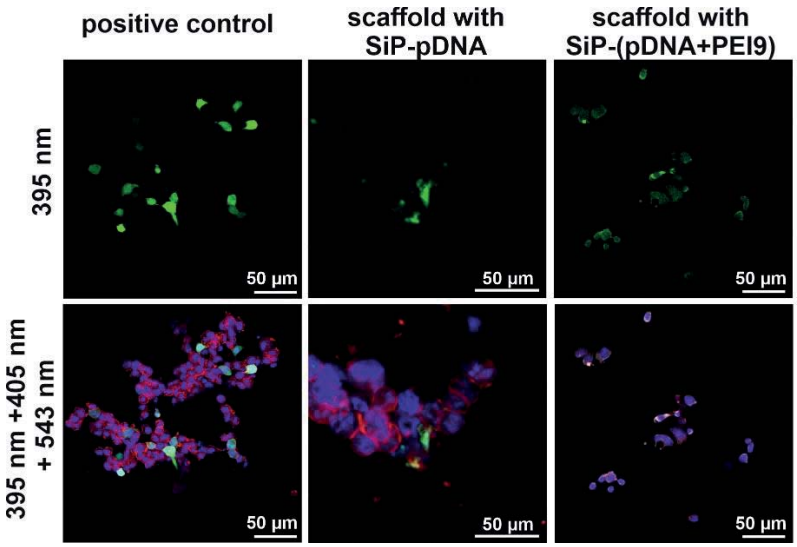


Fig. 1. Fluorescence microscopy data of HEK293T cells cultured on different scaffolds for 5 days. Positive control – cells after pDNA transfection using the TurboFect reagent. Cells are stained with Hoechst 33342 and Phalloidin-TRITC

Conclusion: The results of the work demonstrate that polymer scaffolds made by electrospinning and enriched with pDNA are able to deliver nucleic acids in cells cultivated on this scaffold in order to reprogramming of these cells.

Funding: The study is supported by the Russian State-funded budget project of ICBFM SB RAS No. 121031300042-1.

Список литературы/References

1 Jo J.I., Gao J.Q., Tabata Y. Biomaterial-based delivery systems of nucleic acid for regenerative research and regenerative therapy. *Regener Ther*. 2019;11:123-130

2 Zhuang Y., Cui W. Biomaterial-based delivery of nucleic acids for tissue regeneration. *Adv Drug Delivery Rev*. 2021;176:113885

3 Popova V. et al. Doxorubicin-loaded silica nanocomposites for cancer treatment. *Coatings*. 2023;13:324

doi 10.18699/bgrs2024-9.5-05

Human corneal stromal cells have a high potential for regeneration

Poveshchenko O.V.[1]*, Krasner K.U.[1, 2], Surovtseva M.A.[1], Kim I.I.[1], Bondarenko N.A.[1], Chepeleva E.V.[1], Lykov A.P.[1], Gayvoronskaya A.A.[1], Trunov A.N.[2], Chernykh V.V.[2]

[1] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk Branch of S. Fyodorov Eye Microsurgery Federal State Institution, Novosibirsk, Russia*

* *poveschenkoov@yandex.ru*

Key words: cornea; stromal cells; keratocyte; fibroblast; Relex Smile; corneal regeneration

Motivation and Aim: Corneal blindness is the third leading cause of low vision after cataracts and glaucoma [1]. Due to the shortage of donor corneas, new approaches to the treatment of corneal blindness are being sought. Such approaches are cell therapy and the development of tissue-engineered corneal constructs, which are an alternative to corneal allotransplantation. Corneal stromal cells perform the function of producing and maintaining the extracellular matrix, ensuring morphostructural and biochemical stability, and the transparency of the corneal tissue [2, 3]. The aim of the study is study the functional properties of corneal stromal cells obtained from the source – ReLEx SMILE lenticules and evaluate their regenerative potential *in vivo*.

Methods and Algorithms: Stromal cells were isolated from lenticules. The differentiation of corneal fibroblasts into keratocytes was performed in a selective KGM. The cell production of cytokines and extracellular matrix proteins was determined by ELISA. Injury to the corneal stroma was performed using a sharp 33G needle in the form of a tunnel defect in the thickness of the cornea of C57BL/6 mice. Then stromal cells (fibroblasts ore differentiated keratocytes) were injected into the cornea. Corneal thickness and transparency were assessed using OCT.

Results: Corneal fibroblasts and differentiated keratocytes have functional potential for regeneration, secrete fibronectin, collagen, a number of cytokines, adhesion molecules. It was demonstrated, that injury led to an increase in corneal thickness by 1.3 times on day 15, which was a consequence of trauma – iatrogenic corneal edema. The introduction of fibroblasts and differentiated keratocytes contributed to the restoration of the thickness and transparency of the mouse cornea according to OCT data after 2 months. The corneas of mice that did not receive cell injections after injury were thicker. In addition, this group also revealed high hyperreflectivity of the cornea. The opacity of the cornea of mice in groups without cell treatment after injury was due to the synthesis of collagen III and the presence of myofibroblasts.

Conclusion: Our data indicate, that lenticules, obtained with the ReLEx SMILE, may be a source of highly functional stromal cells for various tasks of regenerative medicine.

References

1. Robaei D., Watson S. Corneal blindness: a global problem. *Clin Exp Ophthalmol*. 2014;42(3):213-214. doi 10.1111/ceo.12330
2. Fukuda K. Corneal fibroblasts: Function and markers. *Exp Eye Res*. 2020;200:108229. doi 10.1016/j.exer.2020.108229
3. Young R.D., Knupp C. et al. Three-dimensional aspects of matrix assembly by cells in the developing cornea. *Proc Natl Acad Sci USA*. 2014;111:687-692. doi 10.1073/pnas.1313561110

9 Симпозиум «Биомедицина, биоинформатика и системная компьютерная биология»

Symposium “Biomedicine, bioinformatics and systems computational biology”

doi 10.18699/bgrs2024-9.6-01

Интерстиций: биоуправление и дренажный перенос жидкостей, кристаллоидов, биополимеров и клеток

Летягин А.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия
letyaginay@bionet.nsc.ru

Ключевые слова: интерстиций; интерстициальная жидкость; дальние экстраваскулярные каналы; дренажные возможности; визуализация; биоуправление

Мотивация и цель: Интерстиций как компонент соединительной ткани диффузно распространяется по всему организму в виде пространственной структуры, контактирует со всеми тканями, обеспечивает движение жидкостей, кристаллоидов, биополимеров и клеток. Кроме этого, интерстициальная жидкость, перемещающаяся через фиброзные матрицы и каналы интерстициальной ткани (= дальние внесосудистые пути) на достаточно большие расстояния, может являться проводником для электромагнитных сигналов различной частоты и интенсивности. Анализ строения и функционирования гликокаликса эндотелиоцитов венозной части капилляров и венул, и лимфататических капилляров и сосудов, данные об изменениях лимфатической системы у глубоководных животных и у людей во время орбитального полета (формирование антравитационной лимфедемы) поставили вопрос о пересмотре схемы Starling'a массопереноса интерстициальной жидкости (ISF) в микроциркуляторном компартменте интерстиция. До недавнего времени был непонятен механизм хорошего эффекта компрессионной терапии лимфедемы, когда регионарные лимфатические сосуды и узлы субтотально гипоплазированы.
Методы и алгоритмы: В научной литературе в последние 10–15 лет получены данные для новой интерпретации морфологической картины интерстиция с помощью технологий: «классическое» инъецирование микрочастиц туши в сочетании с фазовоконтрастной микроскопией и иммуногистохимией; прижизненной флюоресцентной эндостереомикроскопии (pCLM) *in vivo* (в том числе на человеке) с различными флюоресцентными трассерами (флюоресцеин, флуоресцеинизотиоцианат, ферроцианид); сканирующей (SEM) и просвечивающей (TEM) электронной микроскопии; конфокальной лазерной сканирующей микроскопии (CLSM); прижизненной магнитно-резонансной томографии (МРТ) с Gd-содержащими контрастными веществами. 3D-восстановление изображений интерстициальных каналов показали их наличие на локальном (регионарном) и на организменном («дальние экстраваскулярные пути») уровне. Каждый путь имеет три части:
(1) Диффузный (нераспознаваемый) интерстиций с сетеобразными волокнистыми матрицами и межфазной зоной, сецернирующий ISF;
(2) Периневрaльные, паравенозные, параартериальные и перифасциальные дальние пути транспортировки ISF на большие расстояния;

(3) Зона абсорбции ISF в диффузном интерстиции органов, способных утилизировать, метаболизировать или экскретировать во внешнюю среду часть ISF и растворенные в ней метаболиты.
Количество этих работ закономерно увеличивается, в том числе в ведущих мировых журналах (Nature, Scientific Reports). Тем не менее, эти публикации попали под критику: дискуссия на страницах European Journal of Anatomy (официальный журнал Spanish Association of Anatomy), пришла к выводу, что «основное вещество» отождествляется с «тканевой жидкостью» (ISF), в которую погружена фиброзная сетка соединительной ткани, а «отверстия», показанные при прижизненной микроскопии и в криофиксированных образцах, имеют размер до 20–70 мкм [1].
Результаты: Интерстиций, как орган, самый большой в человеческом организме, а объем ISF в интерстиции более чем в три раза превышает объем сердечно-сосудистой и лимфатической систем. Надо также отметить, что внесосудистые потоки в периваскулярных пространствах (PVS) головного мозга были открыты Durand-Fardel, Rudolf Virchow и Charles Robin еще в середине XIX века, а периваскулярный путь вдоль адвентициальной оболочки церебральных вен был идентифицирован как глимфатическая система мозга [2, 3]. Напрашивается вывод, что это прямой аналог интерстициальных каналов в теле человека.
Выводы: Новая интерпретация «легализует» третью дренажную систему организма («дальние экстраваскулярные пути» интерстиция), где ISF движется по системе внесосудистых путей, сформированных коллагено-эластиновыми волоконными матрицами, и частично – за счет мембран фиброцитов. Требуется понимание иерархии физических, химических и биологических механизмов, наличие белков-аналогов аквапорина 4 (AQP4), образующих водопроводящие каналы; энергоэффективность и энергозависимость функционирования каналов интерстиция с биофизической точки зрения.
Финансирование: Исследование поддержано Государственным заданием НИИКЭЛ-филиал ИЦиГ СО РАН (№ FWNR-2022-0012).

Interstitium: biofeedback and drainage transport of fluids, crystalloids, biopolymers and cells

Letyagin A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
letyaginay@bionet.nsc.ru

Key words: interstitium; interstitial fluid; long-distance extravascular channels; drainage capabilities; visualization; biofeedback

Motivation and Aim: The interstitium, as a component of connective tissue, diffuses throughout the body in the form of a spatial structure, contacts all tissues, and ensures the movement of fluids, crystalloids, biopolymers and cells. In addition, interstitial fluid moving through fibrous matrices and channels of interstitial tissue (= long-distance extravascular pathways) over sufficiently long distances can be a conductor for

electromagnetic signals of various frequencies and intensities. Analysis of the structure and functioning of the glycocalyx of endothelial cells of the venous part of capillaries and venules, and lymphatic capillaries and vessels, data on changes in the lymphatic system in deep-sea animals and in humans during orbital flight (formation of gravitational lymphedema) raised the question of revising Starling's scheme of mass transfer of interstitial fluid (ISF) in the microcirculatory compartment of the interstitium. Until recently, the mechanism for the good effect of compression therapy for lymphedema, when regional lymphatic vessels and nodes are subtotally hypoplastic, was unclear.

Methods and Algorithms: In the scientific literature in the last 10–15 years, data have been obtained for a new interpretation of the morphological picture of the interstitium using technologies: "classical" injection of ink microparticles in combination with phase-contrast microscopy and immunohistochemistry; intravital fluorescent endostereomicroscopy (pCLM) *in vivo* (including on humans) with various fluorescent tracers (fluorescein, fluorescein isothiocyanate, ferrocyanide); scanning (SEM) and transmission (TEM) electron microscopy; confocal laser scanning microscopy (CLSM); intravital magnetic resonance imaging (MRI) with Gd-containing contrast agents. 3D reconstruction of images of interstitial channels showed their presence at the local (regional) and organismal ("long-distance extravascular pathways") level. Each pathway has three parts:

(1) Diffuse (unrecognizable) interstitium with network-like fibrous matrices and an interphase zone segregating the ISF;

(2) Perineural, paravenous, paraarterial, and perifascial long-distance transport pathways for ISF;

(3) The absorption zone of ISF in the diffuse interstitium of organs capable of utilizing, metabolizing or excreting part of the ISF and metabolites dissolved in it into the external environment.

The number of these works is naturally increasing, including in the world's leading journals (Nature, Scientific Reports). However, these publications came under criticism: a discussion in the pages of the European Journal of Anatomy (the official journal of the Spanish Association of Anatomy) concluded that the "ground substance" is identified with the "tissue fluid" (ISF), in which the fibrous meshwork is immersed connective tissue, and the "holes" shown during intravital microscopy and in cryofixed samples have a size of up to 20–70 μm [1].

Results: The interstitium, as an organ, is the largest in the human body, and the volume of ISF in the interstitium is more than three times that of the cardiovascular and lymphatic systems. It should also be noted that extravascular flows in the perivascular spaces (PVS) of the brain were discovered by Durand-Fardel, Rudolf Virchow and Charles Robin in the mid-19th century, and the perivascular path along the adventitia of the cerebral veins was identified as the glymphatic system of the brain [2, 3]. The conclusion suggests itself that this is a direct analogue of the interstitial channels in the human body.

Conclusion: The new interpretation "legalizes" the third drainage system of the body ("long-distance extravascular pathways" interstitium), where ISF moves along a system of extravascular pathways formed by collagen-elastin fiber matrices, and partly due to fibrocyte membranes. An understanding of the hierarchy of physical, chemical and biological mechanisms is required, the presence of analogue proteins of aquaporin 4

(AQP4) that form water-conducting channels; energy efficiency and energy dependence of the functioning of interstitial channels from a biophysical point of view.
Funding: The study is supported by the State assignment RICEL – branch of the Institute of Cytology and Genetics SB RAS (No. FWNR-2022-0012).

Список литературы/References

1. Benias P.C. et al. Structure and Distribution of an Unrecognized Interstitium in Human Tissues. *Sci Rep*. 2018;8(1):4947. doi 10.1038/s41598-018-23062-6
2. Iliff J.J. et al. A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β. *Sci Transl Med*. 2012;4(147):147ra111. doi 10.1126/scitranslmed.3003748
3. Rennels M.L., Gregory T.F., Blaumanis O.R. et al. Evidence for a 'Paravascular' fluid circulation in the mammalian central nervous system, provided by the rapid distribution of tracer protein throughout the brain from the subarachnoid space. *Brain Res*. 1985;326(1):47-63. doi 10.1016/0006-8993(85)91383-6

doi 10.18699/bgrs2024-9.6-02

Биомедицинские аспекты влияния мелатонина на экспрессию LYVE1 и HIF-1α в мозге у мышей с генетически детерминированной моделью сахарного диабета II типа

Мичурина С.[1]*, Серых А.[1, 2], Ищенко И.[1], Архипов С.[1], Завьялов Е.[1]

[1] *Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия*

[2] *Федеральный исследовательский центр фундаментальной и трансляционной медицины, СО РАН, Новосибирск, Россия*

* *michurinasv3000@gmail.com*

Ключевые слова: сахарный диабет II типа; глимфатическая система; LYVE1; HIF-1α

Мотивация и цель: В настоящий момент сахарный диабет и осложнения связанные с ним, находятся в десятке лидирующих причин смертности во всем мире. Ввиду того что существующие препараты для лечения диабета имеют многочисленные побочные эффекты, остро стоит необходимость поиска вещества, имеющего высокий терапевтический потенциал. Известно, что течение сахарного диабета II типа (СДII), вследствие существования реципрокной связи мелатонин-инсулин, связано со снижением уровня мелатонина [1]. Кроме того, терапия экзогенным мелатонином, выполненная на клеточных линиях, лабораторных животных и пациентах с СДII показала свою эффективность в облегчении связанных с ним осложнений. Примечательно, что среди таких осложнений существует повышенный риск развития различных неврологических заболеваний, таких как шизофрения и болезнь Альцгеймера, возникающих как закономерное следствие метаболических нарушений в организме, с последующим развитием окислительного и гипоксического стресса в головном мозге [2]. Поэтому мелатонин, являющийся антиоксидантом широкого спектра, и, благодаря своим гидро- и липофильным свойствам легко преодолевающий биологические барьеры представляется нам отличным кандидатом для терапии неврологических осложнений, связанных с СДII. Однако данные о воздействии мелатонина на мозг всё еще противоречивы [3], а потому важно исследовать его воздействие на глимфатическую систему и гипоксию клеток мозга в свете различных моделей СДII. Цель работы − установить влияние мелатонина на экспрессию HIF-1α и рецептора LYVE1 в головном мозге db/db мышей, с генетически обусловленным ожирением и сахарным диабетом II типа.

Методы и алгоритмы: Исследование проводили на трехмесячных самках мышей линии BKS.Cg-Dock7m+/+Leprdb/J (db/db). Животных содержали в барьерных помещениях ЦКП «Центр генетических ресурсов лабораторных животных» ИЦиГ СО РАН (RFMEFI62119X0023). Фенотип данных мышей, ввиду мутации в гене рецептора лептина, проявляется в ожирении, инсулинорезистентности и нарушении метаболизма глюкозы. Были сформированы следующие группы: «Контроль» (n = 5) – интактные животные; «Плацебо» (n = 5) – мыши, ежедневно получавшие внутрижелудочно 200 мкл дистиллированной воды в течение недели;

«Мелатонин» (n = 5) – мыши, которым аналогичным образом вводили мелатонин (MP Biomedical) в дозировке 1 мг/кг массы тела, разведенный в 200 мкл дистиллированной воды. Для исследования забирали головной мозг целиком, совершали стандартную гистологическую пробоподготовку. Парафиновые срезы толщиной 4 мкм получали в соответствии с координатами брегмы в диапазоне от −1.0 до −1.5 мм по Allen Mouse Brain Atlas [4]. Выявление LYVE1 и HIF-1α проводили с помощью непрямого пероксидазного метода при использовании ИГХ-системы детекции Novocastra (RE7110-K), антител специфичных к LYVE1 (ab-14917) и к HIF-1α (AF-1009). С помощью программы "ImageJ" определяли относительную площадь и яркость участков (параметр, обратно пропорциональный концентрации маркера) препаратов мозга, окрашенных на LYVE1 и HIF-1α. С каждого препарата оценивали не менее 20 полей зрения. В качестве целевых анализируемых участков использовали: моторную, соматосенсорную, висцеральную, ретроспленальную, агранулярную островковую, грушевидную коры головного мозга, в зоне поверхностных слоев. Достоверность различий определяли с помощью непараметрического теста Краскалла–Уоллеса, с помощью программного обеспечения STATISTICA 12 (StatSoft, 2015). В качестве исходных данных сравнивали различия по среднему полю зрения в группе. Данные на графиках представлены в виде: медиана (Q1; Q3). Различия считали статистически значимыми при $p < 0.05$.

Результаты: Нами выявлено интенсивное иммуногистохимическое HIF-1α окрашивание на срезах головного мозга db/db мышей, локализированное в цитоплазме нервных и глиальных клеток мозга, особенно в группе «Плацебо». При этом площадь экспрессии HIF-1α у животных, получавших мелатонин, несколько снижалась, по сравнению с группой «Плацебо», также снижалась и концентрация, по сравнению как с контролем, так и с плацебо (рис. 1). Необходимо отметить, что экспрессия HIF-1α была значительно повышена у мышей группы «Плацебо», по сравнению с другими группами животных, что может быть объяснено стрессом, вызванным внутрижелудочным введением мелатонина и плацебо через зонд животным. Это подтверждается данными касательно площади распределения маркера, которая достоверно увеличивалась у мышей группы «Плацебо» и возвращалась к уровню контроля после введения мелатонина. В группе «Контроль» выявлено выраженное LYVE1 окрашивание капилляров и лимфатических эндотелиальных клеток в паренхиме исследованных зон головного мозга. Вокруг капилляров определялись патологически расширенные пространства Вирхова-Робина. Во время вывода животных из эксперимента, в группах «Контроль» и «Плацебо» были обнаружены по одному животному с гидроцефалией. В группе db/db мышей, получавших мелатонин, особей с гидроцефалией отмечено не было. На препаратах наблюдалось уменьшение пространств Вирхова-Робина за счет снижения накопления жидкости в межклеточном пространстве. Морфометрический анализ показал, что у db/db мышей, получавших мелатонин, незначительно снижалась площади экспрессии LYVE1, однако концентрация данного рецептора в глимфатической система мозга остается сравнимой с группами «Контроль» и «Плацебо» (рис. 2).

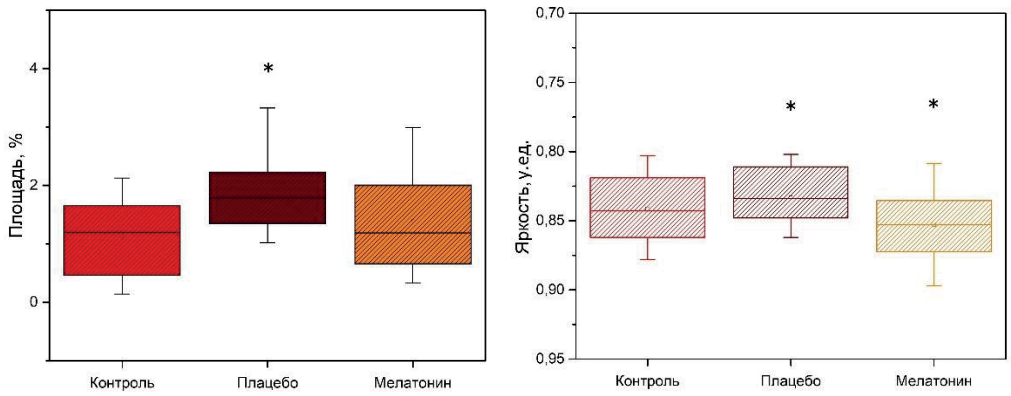


Рис. 1. Площадь и яркость окрашивания HIF-1α в мозге db/db мышей экспериментальных групп: «Контроль», «Плацебо» и «Мелатонин». Примечание: * сравнение с группой «Контроль», * $p < 0.05$, тест Краскалла–Уоллеса

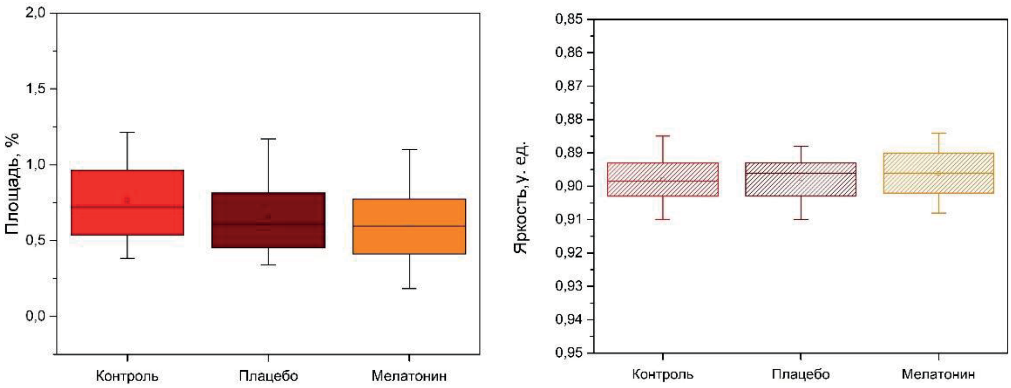


Рис. 2. Площадь и яркость окрашивания рецептора LYVE1 в мозге db/db мышей экспериментальных групп: «Контроль», «Плацебо» и «Мелатонин». Примечание: * сравнение с группой «Контроль», * $p < 0.05$, тест Краскалла–Уоллеса

Выводы: Полученные нами результаты свидетельствуют, что мелатонин оказывает антигипоксический эффект на клетки головного мозга db/db мышей, нормализует жидкостный гомеостаз в глимфатической системе мозга. Важно отметить, что мелатонин хоть и обладает положительным противоотечным действием, но его эффекты обеспечиваются не через регуляцию экспрессии рецептора LYVE1, а посредством других механизмов, вероятно, включающих и биоритмологические, лимфодренажные и антиоксидантные аспекты его влияния на организм.

Финансирование: Исследование выполнено в рамках проекта НИИКЭЛ — филиала ИЦиГ СО РАН (FWNR-2022-0009).

Biomedical aspects of the effect of melatonin on LYVE1 and HIF-1α expression in the brain in mice with a genetically determined model of type II diabetes mellitus

Michurina S.[1]*, Serykh A.[1, 2], Ishchenko I.[1], Arkhipov S.[1], Zavjalov E.[1]

[1] *Research Institute of Clinical and Experimental Lymphology – Branch of Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Federal Research Center of Fundamental and Translational Medicine, SB RAS, Novosibirsk, Russia*

* *michurinasv3000@gmail.com*

Key words: type II diabetes mellitus; glymphatic system; LYVE1; HIF-1α

Motivation and Aim: At present, diabetes mellitus and its complications are among the top ten leading causes of death worldwide. In view of the fact that existing diabetes drugs have numerous side effects, there is an urgent need to find a substance with high therapeutic potential. It is known that the course of type II diabetes mellitus (T2DM), due to the existence of a reciprocal melatonin-insulin relationship, is associated with a decrease in melatonin levels [1]. In addition, exogenous melatonin therapy performed on cell lines, laboratory animals and patients with T2DM has been shown to be effective in alleviating the complications associated with it. Notably, among such complications, there is an increased risk of developing various neurological diseases such as schizophrenia and Alzheimer's disease, arising as a natural consequence of metabolic disorders in the body, with the subsequent development of oxidative and hypoxic stress in the brain [2]. Therefore, melatonin, which is a broad-spectrum antioxidant and, due to its hydro- and lipophilic properties, easily crosses biological barriers, seems to us to be an excellent candidate for the therapy of neurological complications associated with T2DM. However, data on the effects of melatonin on the brain are still contradictory [3], and therefore it is important to investigate its effects on the glymphatic system and brain cell hypoxia in light of different models of T2DM. The aim of this work was to determine the effect of melatonin on the expression of HIF-1α and LYVE1 receptor in the brain of db/db mice, with genetically determined obesity and type II diabetes mellitus.

Methods and Algorithms: The study was performed on three-month-old female BKS.Cg-Dock7m+/+Leprdb/J (db/db) mice. Animals were kept in barrier rooms of the collective use center "Laboratory Animal Genetic Resources Center" IC&G SB RAS (RFMEFI62119X0023). The phenotype of these mice, due to a mutation in the leptin receptor gene, is manifested by obesity, insulin resistance and impaired glucose metabolism. The following groups were formed: "Control" (n = 5) – intact animals; "Placebo" (n = 5) – mice that received intragastrically 200 μl of distilled water daily for a week; "Melatonin" (n = 5) – mice that were similarly administered melatonin (MP Biomedical) at a dosage of 1 mg/kg body weight diluted in 200 μl of distilled water. The whole brain was taken for the study and standard histologic sample preparation was performed. Paraffin sections 4 μm thick were obtained according to bregma coordinates ranging from –1.0 to –1.5 mm using the Allen Mouse Brain Atlas [4]. LYVE1 and HIF-1α were detected by indirect peroxidase method using Novocastra IHC detection system (RE7110-K), antibodies specific to LYVE1 (ab-14917) and to HIF-1α (AF-1009). The relative area and brightness of areas (a parameter inversely proportional to marker concentration) of brain preparations stained for LYVE1 and HIF-1α were determined

using the ImageJ program. At least 20 fields of view were evaluated from each preparation. As target analyzed areas were used: motor, somatosensory, visceral, retrosplenial, agranular insular, piriform cortex, in the zone of superficial layers. The significance of differences was determined using the nonparametric Kruskal-Wallis test, using STATISTICA 12 software (StatSoft, 2015). The differences in mean visual field in the group were compared as baseline data. Data on the graphs are presented as: median (Q1; Q3). Differences were considered statistically significant at $p < 0.05$.

Results: We found intense immunohistochemical HIF-1α staining on brain slices of db/db mice localized in the cytoplasm of nerve and glial cells in the brain, especially in the placebo group. The area of HIF-1α expression was slightly decreased in animals receiving melatonin compared to the placebo group, and the concentration was also decreased compared to both control and placebo (Fig. 1).

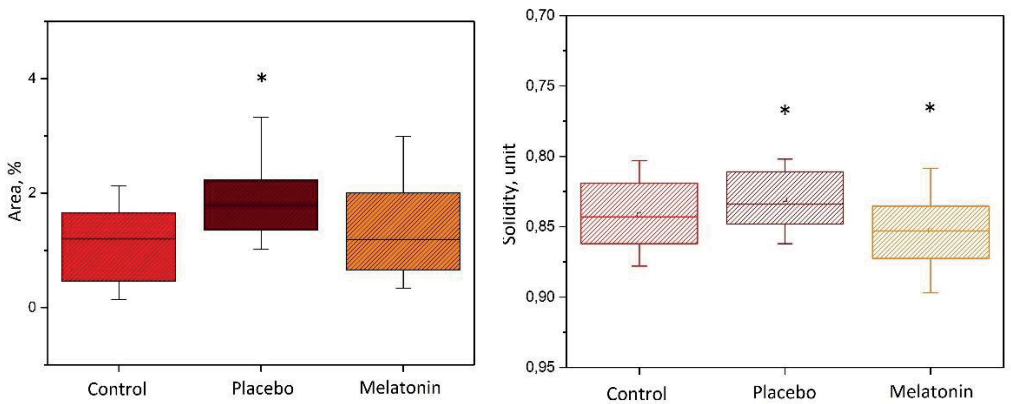


Fig. 1. Area and brightness of HIF-1α staining in the brains of db/db mice of experimental groups: "Control", "Placebo" and "Melatonin". Note: * comparison with control group, * $p < 0.05$, Kruskal–Wallis test

It should be noted that the expression of HIF-1α was significantly upregulated in mice of the placebo group, compared with the other groups of animals, which may be explained by the stress induced by intragastric administration of melatonin and placebo through the probe to the animals. This is confirmed by the data concerning the distribution area of the marker, which significantly increased in mice of the placebo group and returned to the control level after melatonin administration. The control group showed pronounced LYVE1 staining of capillaries and lymphatic endothelial cells in the parenchyma of the studied brain areas. Pathologically dilated Virchow-Robin spaces were defined around capillaries. At the time of withdrawal of animals from the experiment, one animal with hydrocephalus each was found in the control and placebo groups. In the db/db group of mice receiving melatonin, no individuals with hydrocephalus were observed. The preparations showed a decrease in Virchow-Robin spaces due to decreased fluid accumulation in the intercellular space. Morphometric analysis showed that in db/db mice treated with melatonin, the expression area of LYVE1 was slightly decreased; however, the concentration of this receptor in the brain glymphatic system remains comparable to the control and placebo groups (Fig. 2).

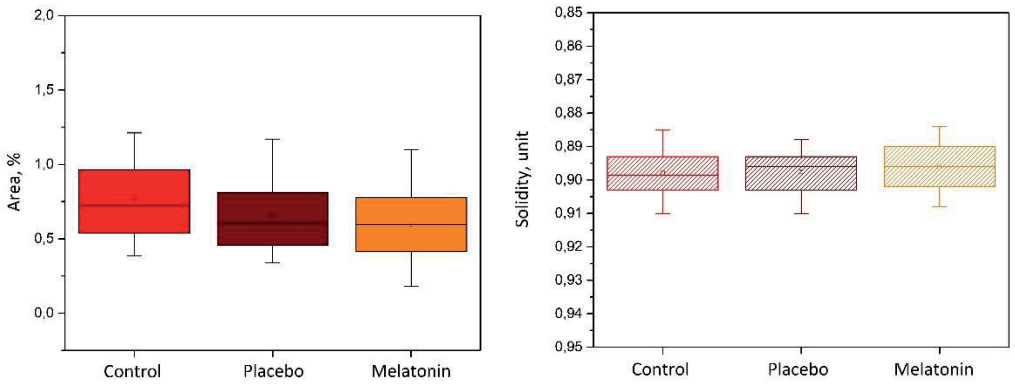


Fig. 2. Area and brightness of LYVE1 receptor staining in the brains of db/db mice of experimental groups: “Control”, “Placebo” and “Melatonin”. Note: * comparison with control group, * $p < 0.05$, Kruskal–Wallis test

Conclusion: Our results indicate that melatonin has an antihypoxic effect on brain cells of db/db mice and normalizes fluid homeostasis in the brain glymphatic system. It is important to note that although melatonin has a positive antiedematous effect, its effects are provided not through regulation of LYVE1 receptor expression, but through other mechanisms, probably including biorhythmological, lymphatic drainage and antioxidant aspects of its influence on the organism.

Funduing: The study was carried out within the framework of the project of Research Institute of Clinical and Experimental Lymphology – Branch of Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (FWNR-2022-0009).

Список литературы/References

1. Patel R., Parmar N. et al. Diabetes mellitus and melatonin: Where are we? *Biochimie*. 2022;202:2-14. doi 10.1016/j.biochi.2022.01.001
2. Shen S., Liao Q., Wong Y. et al. The role of melatonin in the treatment of type 2 diabetes mellitus and Alzheimer's disease. *Int J Biol Sci*. 2022;18(3):983-994. doi 10.7150/ijbs.66871
3. Garaulet M., Qian J., Florez J. et al. Melatonin effects on glucose metabolism: Time to unlock the controversy. *Trends Endocrinol Metab*. 2020:31(3):192-204. doi 10.1016/j.tem.2019.11.011
4. Reference atlas: Allen mouse brain atlas. URL: https://mouse.brain-map.org (дата обращения: 12.02.2024)

doi 10.18699/bgrs2024-9.6-03

Фотобиомодуляция и интерстиций в медицине: новые вызовы и перспективы

Небрат В.
Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия
academvlad@yandex.ru

Ключевые слова: фотобиомодуляция; жидкости интерстиция; структуры воды; модель человека

Мотивация и цель: Фотобиомодуляция (ФБМ), это способ воздействия на организм человека светом, использующий неионизирующие источники фотонов, включая лазеры, светодиоды и широкополосный свет в видимом и инфракрасном диапазоне спектра [1]. ФБМ применяют в течение последних 30 лет во многих областях медицины для терапии разнообразных заболеваний и патологических состояний организма человека, в том числе при заживлении ран, болях при артрите, остром инсульте и других [2]. Эффективность ФБМ терапии определяют множество факторов, среди которых наиболее значимыми являются оптические параметры кожи и тканей организма и технические характеристики используемых источников фотонов [3]. Роль этих факторов выясняют в каждом конкретном случае применения ФБМ терапии с помощью исследований, как теоретических, так и экспериментальных. Иногда они позволяют выявить важные закономерности и повысить эффективность ФБМ терапии. Но в большинстве клинических случаев такие исследования провести невозможно и эффективность определяют практические знания врача. Однако эти знания оказываются бесполезными, когда врачи сталкиваются с пациентами, которые не реагируют на ФБМ терапию [4]. Такая ситуация в клинической медицине обусловлена тем, что механизм лечебного действия ФБМ до сих пор остается мало изученным и плохо понимаемым. Этому пониманию также мало способствуют предлагаемые идеи и гипотезы механизмов ФБМ, которые рассматривают в качестве первичных мишеней фотонов, прежде всего отдельные клетки внутри организма [5]. Идея автора о том, что такими мишенями могут служить интерстициальные жидкости (ISF), которые присутствуют во всех тканях тела человека, является новой гипотезой ФБМ терапии. Целью данной работы является изложение и обоснование гипотезы о роли ISF в ФБМ терапии.
Методы и алгоритмы: Подвергали анализу ISF тела человека. Для этого изучали эксперименты по визуализации каналов ISF тела человека с помощью метода МРТ с применением широко используемого индикатора Gd-DTPA и метода гистологического исследования [6]. Затем, изучали ISF тела человека с помощью моделирования, используя ранее разработанную физико-математическую модель водного строения человека [7]. Для этого представляли структуру сети каналов ISF тела в виде графа из циклически взаимосвязанных между собой элементов. При этом в качестве элементов графа и их связей рассматривали каналы ISF тела. Граф интерпретировали согласно модели как систему циркуляции воды ISF организма, а возникновение и функционирование элементов сети объясняли процессами самоорганизации воды ISF тела, постоянно и непрерывно

происходящими под действием фотонов, поступающих на землю от солнца. Рассматривали их действие, как природный фотонный механизм самоорганизации сети водных каналов ISF тела, единый для всех элементов сети. Полагали, что данный механизм может играть важную роль в организме и оказывать существенное влияние на функциональные состояния человека. Поэтому анализировали работу ISF сети графа и выясняли закономерности природного фотонного механизма и его роль в организме. Найденные закономерности считали общими для природных фотонов и фотонов ФБМ терапии. Далее оценивали влияние параметров и характеристик источников фотонов ФБМ терапии на эти закономерности. Последствия этого влияния интерпретировали как воздействие на функциональные состояния человека, как нормальные, так и патологические. Полученные результаты использовали для теоретического обоснования гипотезы о роли ISF в ФБМ терапии и объяснения данных клинических экспериментов ее эффективного применения.

Результаты: Сформулирована гипотеза о роли ISF в ФБМ терапии, которая гласит, что вода ISF организма является первичной мишенью и главным участником фотобиологических процессов, которыми управляют природные фотоны, поступающие на землю от солнца, а при нарушении управления, эту роль могут выполнять фотоны ФБМ терапии. Предоставлено обоснование гипотезы, основанное на анализе экспериментальных данных по визуализации каналов ISF с помощью МРТ метода и гистологических исследований, а также результатов моделирования, использующих ранее разработанную физико-математическую модель водного строения человека. Анализ экспериментальных данных показал, что в теле человека действительно существует сеть каналов переноса потока ISF на большие расстояния за пределами сосудистой циркуляции. При этом точно установлено, что существует как минимум четыре типа каналов участвующих в переносе потоков ISF вдоль кожи, артерий, вен и нервов [8]. Эти результаты ясно демонстрируют, что в теле человека реально существуют каналы ISF и то, что кожный путь потока ISF, основным компонентом которого является вода, может являться первичной мишенью фотонов. Этот вывод также подтверждают данные о содержании воды в коже (более 70 %) и пики повышенного поглощения воды в диапазоне длин волн фотонов ФБМ терапии: 740, 835, 970, 1180, 1650, 3400 нм, которые наблюдаются в эпидермисе, дерме и подкожно-жировой клетчатке [9]. Анализ результатов моделирования позволил установить важные закономерности действия природных фотонов на процессы самоорганизации воды ISF организма и выразить их следующим образом. После поглощения фотонов водой ISF происходит диссоциация молекул воды на ионы H+ и OH-. Затем направленное движение этих заряженных ионов в водной среде ISF организма создает ионные (электрические) токи двух полярностей, циркулирующие по водным каналам сети ISF одновременно в двух противоположных направлениях. Это приводит к постоянному изменению направления суммарного тока сети ISF в зависимости от текущего соотношения вкладов двух ионных токов и одновременно к возбуждению вокруг водных каналов переменного электрического и магнитного поля. Данный вывод четко показывает, что вода ISF организма может выполнять роль не только первичной мишени, но и являться участником разнообразных электромагнитных процессов организма, включая и фотобиологические.

Выводы: Главный вывод проведенного исследования состоит в том, что проблема эффективности применения ФБМ терапии в клинической медицине связана с

непониманием физической роли природных фотонов и воды организма. Вода, которую принято рассматривать в медицине только как универсальный растворитель, является согласно современным физическим представлениям активной упорядоченной средой организма. Поэтому новым вызовом для ФБМ медицины является необходимость использования физического взгляда на процессы организма, а не только клеточного, доминирующего в медицине. Эта необходимость обусловлена тем, что именно фундаментальные законы физики определяют закономерности взаимодействия в организме фотонов и воды ISF, которая является основным компонентом всех тканей тела человека. О важной роли учета этих законов в ФБМ терапии свидетельствуют эффективные результаты экспериментальных исследований по лечению сложных состояний, например таких как, регенерация целого органа, что недоступно современной медицине [10]. Эти и другие аналогичные результаты показывают, что исследования физических взаимодействий фотонов и воды ISF организма, которым до настоящего времени уделялось ограниченное внимание, открывают новые перспективы применения ФБМ терапии в клинической медицине. Вместе с этим, эти исследования позволят выяснить сложные и, по-видимому, разнообразные функции сети каналов ISF в организме человека.

Photobiomodulation and interstitium in medicine: new challenges and prospects

Nebrat V.*

Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

** academvlad@yandex.ru*

Key words: photobiomodulation; interstitial fluid; water structures; human model

Motivation and Aim: Photobiomodulation (PBM) is a method of exposing the human body to light using non-ionizing photon sources, including lasers, LEDs and broadband light in the visible and infrared range of the spectrum [1]. PBM has been used over the past 30 years in many areas of medicine for the treatment of various diseases and pathological conditions of the human body, including wound healing, arthritis pain, acute stroke, and others [2]. The effectiveness of PBM therapy is determined by many factors, among which the most significant are the optical parameters of the skin and body tissues and the technical characteristics of the photon sources used [3]. The role of these factors is clarified in each specific case of the use of PBM therapy through research, both theoretical and experimental. Sometimes they make it possible to identify important patterns and increase the effectiveness of PBM therapy. But in most clinical cases, such studies cannot be carried out and the effectiveness is determined by the practical knowledge of the doctor. However, this knowledge is of no use when clinicians encounter patients who do not respond to PBM therapy [4]. This situation in clinical medicine is due to the fact that the mechanism of the therapeutic action of PBM still remains poorly studied and poorly understood. This understanding is also of little assistance from the proposed ideas and hypotheses about the mechanisms of PBM, which are considered, first of all, individual cells of the body as primary targets for photons [5]. The author's idea that such targets could be interstitial fluids (ISF), which

are present in all tissues of the human body, is a new hypothesis for PBM therapy. The purpose of this work is to present and substantiate the hypothesis about the role of ISF in PBM therapy.

Methods and Algorithms: The ISF of the human body was studied and analyzed. To do this, we studied experiments on imaging the ISF channels of the human body using the MRI method with the application of the widely used indicator Gd-DTPA and the method of histological examination [6]. Then the human ISF was studied using a previously developed physico-mathematical model of the human water system [7]. To do this, we represented the structure of the ISF channel network of the body in the form of a graph of cyclically interconnected elements. In this case, the ISF channels of the body were considered as elements of the graph and their connections. The graph was interpreted according to the model as the body's ISF water circulation system, and the emergence and functioning of the network elements was explained by the processes of self-organization of the body's ISF water, constantly and continuously occurring under the influence of photons arriving on the earth from the sun. Their action was considered as a natural photonic mechanism of self-organization of the network of water channels of the ISF body, common for all elements of the network. It was believed that this mechanism could play an important role in the body and have a significant impact on the functional state of a person. Therefore, we analyzed the work of the ISF graph network and found out the patterns of the natural photonic mechanism and its role in the body. The discovered patterns were assumed to be common for natural photons and photons of PBM therapy. Next, we assessed the influence of the parameters and characteristics of photon sources of PBM therapy on these patterns. The consequences of this influence have been interpreted as affecting human functional states, both normal and pathological. The results obtained were used to theoretically substantiate the hypothesis about the role of ISF in PBM therapy and to explain the data from clinical experiments of its effective use.

Results: A hypothesis has been formulated about the role of ISF in PBM therapy, which states that the ISF water body is the primary target and main participant in photobiological processes controlled by natural photons coming to the earth from the sun, and if control is disrupted, this role can be played by photons of PBM therapy. A substantiation of the hypothesis is provided, based on the analysis of experimental data on the visualization of ISF channels using the MRI method and histological studies, as well as modeling results using a previously developed physico-mathematical model of the human water system. Analysis of experimental data has shown that in the human body there really is a network of channels that transport ISF flow over long distances outside the vascular circulation. It has been precisely established that there are at least four types of channels involved in the transfer of ISF flows along the skin, arteries, veins and nerves [8]. These results clearly demonstrate that ISF channels actually exist in the human body and that the cutaneous ISF pathway, of which water is the main component, may be the primary target of photons. This conclusion is also confirmed by data on the water content in the skin (more than 70 %) and peaks of increased water absorption in the wavelength range of photons of PMB therapy: 740, 835, 970, 1180, 1650, 3400 nm, which are observed in the epidermis, dermis and subcutaneous fat [9]. Analysis of the modeling results made it possible to establish important patterns of the action of natural photons on the processes of self-organization of ISF water in the body and express them as follows. After photons are absorbed by ISF water, the water molecules dissociate into H+ and OH-ions. Then the directed movement of these charged ions in the water

environment of the ISF of the body creates ionic (electric) currents of two polarities, circulating through the water channels of the ISF network simultaneously in two opposite directions. This leads to a constant change in the direction of the total current of the ISF network, depending on the current ratio of the contributions of the two ion currents, and at the same time, to the excitation of an alternating electric and magnetic field around the water channels. This conclusion clearly shows that ISF water in the body can serve not only as a primary target, but also as a participant in a variety of electromagnetic processes in the body, including photobiological ones.

Conclusions: The main conclusion of the study is that the problem of the effectiveness of the use of PBM therapy in clinical medicine is associated with a lack of understanding of the physical role of natural photons and body water. Water, which is usually considered in medicine only as a universal solvent, is, according to modern physical concepts, an active, ordered medium of the body. Therefore, a new challenge for PBM medicine is the need to use a physical view of the processes of the body, and not just the cellular one, which dominates in medicine. This need is due to the fact that it is the fundamental laws of physics that determine the patterns of interaction in the body of photons and ISF water, which is the main component of all tissues of the human body. The importance of taking these laws into account in the treatment of PBM is evidenced by the effective results of experimental studies on the treatment of complex conditions, such as the regeneration of an entire organ, which are inaccessible to modern medicine [10]. These and other similar results indicate that studies of the physical interactions of photons and body ISF water, which have received limited attention to date, open up new prospects for the application of PBM therapy in clinical medicine. At the same time, these studies will simultaneously elucidate the complex and apparently diverse functions of the ISF channel network in the human body.

Список литературы/References

1. Heiskanen V. et al. Photobiomodulation: lasers vs light emitting diodes? *Photochem Photobiol Sci*. 2018;17(8):1003-1017
2. Yang M. et al. Current application and future directions of photobiomodulation in central nervous diseases. *Neural Regen Res.* 2021;16(6):1177-1185
3. Zein R. et al. Review of light parameters and photobiomodulation efficacy: dive into complexity. *J Biomed Optics*. 2018;23(12):120901
4. Gavish L. et al. Microcirculatory response to photobiomodulation—why some respond and others do not: a randomized controlled study. *Lasers Surg Med.* 2020;52(9):863-872. doi 10.1002/lsm.23225
5. de Freitas L.F., Hamblin M.R. Proposed mechanisms of photobiomodulation or low-level light therapy. *IEEE J Sel Top Quantum Electron*. 2016;22(3):7000417
6. Li H. et al. An extravascular fluid transport system based on structural framework of fibrous connective tissues in human body. *Cell Prolif.* 2019;52(5):e12667. doi 10.1111/cpr.12667
7. Nebrat V. Human water model: interstitium and meridians of traditional Chinese medicine. In: Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022). Novosibirsk, 2022;816. doi 10.18699/SBB-2022-474
8. Li H. et al. A long-distance fluid transport pathway within fibrous connective tissues in patients with ankle edema. *Clin Hemorheol Microcirc*. 2016;63:411-421
9. Setchfield K. et al. Relevance and utility of the in-vivo and ex-vivo optical properties of the skin reported in the literature: a review. *Biomed Opt Express*. 2023;14(7):3555-3583
10. Cha Y., Choe H. et al., Treatment of patient with cirrhosis-ascites by the mineral pulse light stimuli on LV acupoints – a case report. *Res Square*. 2020. doi 10.21203/rs.3.rs-52673/v1

doi 10.18699/bgrs2024-9.6-04

Биоподобие и биосовместимость в физико-химических свойствах медицинских сорбентов

Рачковская Л.Н.*, Летягин А.Ю., Мичурина С.В., Бгатова Н.П., Рахимова Н.Н., Королев М.А.

Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия
** noolit@niikel.ru*

Ключевые слова: медицинские сорбенты; биоподобие; биосовместимость; безопасность; детоксикация

В современных медицинских технологиях важно понимание тонких биофизических механизмов санации внутренней среды организма, так как микроокружение определяет структуру и функцию клеток и органов [1]. Превышение допустимого уровня токсических продуктов экзо- и эндогенного происхождения во внутренней среде организма является причиной патологических состояний. Сорбционные материалы и биомедицинские технологии их применения обеспечивают детоксикацию внутренней среды организма с помощью гемосорбции или путем энтерального или аппликационного их применения. Сорбенты имеют большую величину удельной поверхности за счет «набора» различающихся по размерам пор (нанопор), сопоставимых с размерами удаляемых токсических агентов. Роль химической природы поверхности сорбента также велика, она должна быть, прежде всего, ареактивной в биологическом плане, совместимой с биологическими тканями, то есть гидрофильно-гидрофобной. Механизм взаимодействия пористого контента с биологическими тканями еще недостаточно изучен. Однако не вызывает сомнения, что саногенное действие сорбентов определяется их способностью связывать и выводить из организма токсические продукты из мест их расположения – из крови, лимфы, с раневых поверхностей кожи и ЖКТ. Известен феномен сорбционной детоксикации тканей, в основе которого лежит извлечение токсических метаболитов, микробных клеток, токсинов из очага поражения (воспалительный фокус, ожог, травмированные ткани). С другой стороны, являясь химически, казалось бы, индифферентным, сорбент влияет на физиологические константы организма, не связанные напрямую с процессом энтеросорбции, гемосорбции или аппликационной сорбции. Сорбент выступает как пусковой механизм каскада биохимических реакций от местного до организменного уровня. Как синергист лимфатической системы сорбент исполняет роль лимфопротектора, лимфокорректора. Найдена схожесть действия гранул сорбента с лимфатическим узлом, который выполняет лимфодетоксикацию на основе лимфодренажной функции. Было показано, что сорбент исполняет при лечении ран роль временного «костыльного» (искусственного) лимфоузла, поглощая токсины и протезируя лимфодренажную функцию. Сорбент защищает мембраны клеток от воздействия токсинов, сорбируя их на своей поверхности, тем самым обеспечивает и восстанавливает нормальное функционирование при нарушении биохимических процессов внутри клетки. Тем самым сорбент способствует восстановлению

саморегуляции на уровне клеточных органелл, прежде всего – митохондрий. Таким образом, сорбент разрывает патологические цепочки и круги, восстанавливая реципрокность на уровне клетки и ткани. Более того, сорбционные свойства и природу поверхности пор сорбента можно модифицировать активными молекулами – как биополимерами с биоактивными свойствами и большими молекулами (антибиотики, сульфаниламиды, гормоны), так и микроэлементами (литий, серебро). При этом изменяется механизм действия адсорбированных молекул, изменяются количественные (концентрационные) показатели эффективных дозировок. Этот эффект потенцирования биохимических фармакодинамических процессов еще требует изучения. Сорбент, извлекая «избыточную» жидкость, дегидратирует ткани до нормального уровня, параллельно обеспечивая специфическую адсорбцию кислых метаболитов, что приводит к повышению pH в очаге поражения и благоприятно сказывается на течении патологического процесса, на повышение эффективности лекарственных средств, на стимуляцию репаративных процессов на клеточном и тканевом уровнях.

Финансирование: Исследование поддержано Государственным заданием НИИКЭЛ – филиал ИЦи Г СО РАН (№ FWNR-2022-0009).

Biosimilarity and biocompatibility in the phisico-chemical properties of medical sorbents

Rachkovskaya L.N.*, Letyagin A.Yu., Michurina S.V., Bgatova N.P.,
Rakhimova N.N., Korolev M.A.

Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

** noolit@niikel.ru*

Key words: medical sorbents; biosimilarity; biocompatibility; safety; detoxication

In modern medical technologies, it is important to understand the subtle biophysical mechanisms of sanitation of the internal environment of the body, since the microenvironment determines the structure and function of cells and organs [1]. Exceeding the permissible level of toxic products of exogenous and endogenous origin in the internal environment of the body is the cause of pathological conditions. Sorption materials and biomedical technologies of their application provide detoxification of the internal environment of the body by hemosorption or by enteric or applicative application. Sorbents have a large specific surface area due to a "set" of pores (nanopores) differing in size, comparable to the size of the toxic agents being removed. The role of the chemical nature of the sorbent surface is also great, it must first of all be biologically active, compatible with biological tissues, that is, hydrophilic-hydrophobic. The mechanism of interaction of porous content with biological tissues has not yet been sufficiently studied. However, there is no doubt that the sanogenic effect of sorbents is determined by their ability to bind and remove toxic products from the body from their locations – from blood, lymph, from wound surfaces of the skin and gastrointestinal tract. The phenomenon of sorption detoxification of tissues is known, which is based on the extraction of toxic metabolites, microbial cells, toxins from the lesion (inflammatory

focus, burn, injured tissues). On the other hand, being chemically seemingly indifferent, the sorbent affects the physiological constants of the body that are not directly related to the process of enterosorption, hemosorption or application sorption. The sorbent acts as a trigger for a cascade of biochemical reactions from the local to the organizational level. As a synergist of the lymphatic system, the sorbent acts as a lymphoprotector, lymphocorrector. The similarity of the action of sorbent granules with the lymph node, which performs lymph detoxification based on lymphatic drainage function, was found. It has been shown that the sorbent performs the role of a temporary "crutch" (artificial) lymph node in the treatment of wounds, absorbing toxins and prosthetics of lymphatic drainage function. The sorbent protects cell membranes from the effects of toxins by sorbtion them on its surface, thereby ensuring and restoring normal functioning in case of violation of biochemical processes inside the cell. Thus, the sorbent promotes the restoration of self-regulation at the level of cellular organelles, primarily mitochondria. Thus, the sorbent breaks pathological chains and circles, restoring reciprocity at the cell and tissue levels. Moreover, the sorption properties and the nature of the sorbent's pore surface can be modified by active molecules – both biopolymers with bioactive properties and large molecules (antibiotics, sulfonamides, hormones) and trace elements (lithium, silver). At the same time, the mechanism of action of the adsorbed molecules changes, and the quantitative (concentration) indicators of effective dosages change. This effect of potentiation of biochemical pharmacodynamic processes still needs to be studied. The sorbent, extracting the "excess" liquid, dehydrates tissues to a normal level, simultaneously providing specific adsorption of acidic metabolites, which leads to an increase in pH in the lesion and has a beneficial effect on the course of the pathological process, on increasing the effectiveness of drugs, on stimulating reparative processes at the cellular and tissue levels.

Funding: The study is supported by the State assignment RICEL – Branch of the Institute of Cytology and Genetics, SB RAS (No. FWNR-2022-0009).

Список литературы/References

1. Konenkov V.I., Borodin Yu.I., Lyubarsky M.S. Lymphology. Russian Academy of medical Sciences, Siberian branch, Research Institute of clinical and experimental lymphology. Novosibirsk: Manuscript, 2012

doi 10.18699/bgrs2024-9.6-05

Перспективы использования достижений физики жидких кристаллов в биологии

Трашкеев С.[1, 2]*, Стаценко П.[1], Хомяков М.[1], Швецов С.[3]

[1] *Институт лазерной физики СО РАН, Новосибирск, Россия*

[2] *Институт химической кинетики и горения СО РАН, Новосибирск, Россия*

[3] *Ереванский государственный университет, Ереван, Армения*

* *sitrskv@mail.ru*

Ключевые слова: жидкий кристалл; ориентационная механика; дисклинация; тепловые структуры

Мотивация и цель: Планируется рассмотреть различные эффекты в жидких кристаллах (ЖК), которые обнаруживают аналогии с биологическими процессами [1–6]. В частности, в работе [6] предлагается ввести термин «биологические жидкие кристаллы». Более конкретно, в работе предлагается ряд подходов для исследования жидкокристаллических структурных объектов в микробиологии и других областях науки о живом.

В докладе приведены исследования жидкокристаллических структур, образованных под действием формы занимаемого или вытесненного объема, внешних электромагнитных полей, включая когерентные, тепловых потоков и ионизирующего излучения.

Жидкий кристалл (ЖК) – среда с дополнительными внутренними степенями свободы, определяемыми ориентацией длинных осей, взаимодействующих между собой молекул. Ориентационное упорядочивание может изменяться относительно слабыми внешними воздействиями. Это свойство приводит к представлению жидкокристаллической среды как мягкой или структурированной материи. Анизотропия физических свойств ЖК приводит с существенному усложнению механики и электродинамики анизотропных жидкостей и, как следствие, появлению многих свойств, не имеющих аналогов в обычных веществах.

Интерес к изучению подобных сред определяется современными тенденциями в изучении частично упорядоченных сред. В частности, рассматриваются аналогии между свойствами ЖК и их композитов с биологическими объектами. К научным и прикладным аспектам можно отнести принципы управления движением локализованных структур и включений в ЖК среде, реализуемых с помощью внешних полей или дистанционно, с использованием когерентного излучения. Механизмы динамического управления структурами в ЖК могут быть перенесены на биологические среды с включением микро- и наночастиц.

Методы и алгоритмы: В качестве методов исследования в основном рассматриваются оригинальные приемы и подходы, разработанные в ИЛФ СО РАН и других лабораториях (ФИРАН, Москва; ЕГУ, Ереван), которые, на наш взгляд, могут быть применены для изучения биологических объектов.

Из достаточно хорошо известных эффектов в ЖК следует отметить переход Фредерикса в постоянных магнитном и электрическом полях [1], более сложные ориентационные динамические эффекты [7] и их обобщение в виде светоиндуцированной переориентации [8]. Но более перспективными могут оказаться эффекты, при которых происходит образование дефектов, особых точек,

в которых ориентация ЖК не определена [4]. Источниками могут служить нанообъекты, такие как наночастицы, молекулы полимеров [9] и даже атомарные структуры от ионизирующего излучения [10]. Не менее интересными представляются процессы самоорганизации нано- и микрообъектов за счет регулярного расположения дефектов [11] в периодических электрических полях [4] или при воздействии лазерного излучения. Кроме механических включений, создающих дефект, предлагаются дистанционные методы генерации и управления структурами ЖК с использованием когерентных источников [4, 12]. Механизмом появления таких образований является обнаруженный термоориентационный эффект [13], обязанный переориентацией молекул ЖК тепловыми и гидродинамическими потоками. В работе рассмотрен ряд тепловых и ориентационных структур, появляющихся в поглощающих ЖК под действием локализованного излучения.

В качестве аналогий с процессами, происходящими в биологических системах, можно привести примеры из гидродинамики жидких кристаллов, являющиеся достаточно сложными явлениями из-за необходимости учета структурных особенностей анизотропной жидкости [14]. В частности, стоит отметить эффект Лемана, приводящий к вращению молекул холестерического ЖК под действием градиента температуры, а также явления, связанные с течением холестерика под действием слабых градиентов давления, приводящие к эффектам «просачивания» с резким увеличением эффективной вязкости и появлению азимутальных компонент скорости.

Для интерпретации эффектов, приводимых в работе, рассматривается теоретическая модель, основанная на представлении ЖК в виде распределенной среды [15]. Основные положения предлагаемого подхода позволяют сделать вывод о возможности описания жидкокристаллического состояния среды в электромагнитных полях на наноразмерных масштабах, включая особые точки ориентации и фазовых переходов в деформированных структурах [15, 16].

Результаты: Результатом настоящей работы могут служить перечисленные свойства жидкокристаллических сред и их композитов, рассматриваемые с точки зрения некоторой аналогии с живыми объектами и предложения более глубокого исследования этих аналогий, как некоторых биологических жидких кристаллов.

Выводы: В качестве заключения приводятся предполагаемые авторами направления для междисциплинарных исследований.

1. Визуализация в ЖК нано- и микрообъектов, включая биологические молекулы, а также визуализация микроструктур на поверхности.
2. Определение характера поверхностного взаимодействия биологических микрообъектов при контакте с ЖК.
3. Исследование характера взаимодействия между биологическими микрообъектами.

Возможные механизмы:

а) Гидродинамический импульс с образованием канала передачи.

б) Упругая деформация без течения.

в) Электрический импульс, образующийся вследствие флексоэлектрического эффекта.

г) Изменение параметра порядка. Энтропийная (?) форма.

д) Комбинации предыдущих механизмов.

4. Управление микроразмерными биологическими структурами (образование, перемещение, вращение и т. п.) на основе приемов, разработанных для ЖК с использованием электрических полей, когерентного излучения (лазерный пинцет), акустики, тепловых потоков и (или) градиентов параметра порядка.
5. Восстановление (реконструкция) формы объекта при его визуализации (п. 1) по ориентационному отклику ЖК. Определение степени корректности для обратной задачи.
6. Адаптация и доработка теоретических моделей ЖК для описания частично-упорядоченных биологических сред при наличии внешнего воздействия.

В заключение отметим, что эффекты в жидкокристаллических структурах или их аналогии могут быть эффективно использованы для исследований области биологии, медицины и других смежных науках.

Финасирование: Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 годы), регистрационный номер темы 1021062211014-7-1.3.6.

Exploring the potential of liquid crystal physics in biology

Trashkeev S.[1, 2]*, Statsenko P.[1], Khomyakov M.[1], Shvetsov S.[3]

[1] *Institute of Laser Physics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia*

[2] *Institute of Chemical Kinetics and Combustion, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia*

[3] *Yerevan State University, Yerevan, Armenia*

* *sitrskv@mail.ru*

Key words: liquid crystal; orientational mechanics; disclination; thermal structures

Motivation and Aim: Various effects in liquid crystals (LCs) that exhibit analogies with biological processes [1–6] are planned to be considered. In particular, the term "biological liquid crystals" is proposed in the work [6]. More specifically, the work proposes a range of approaches for studying liquid crystalline structural objects in microbiology and other areas of life science.

The report presents research on liquid crystalline structures formed under the influence of the shape of the occupied or displaced volume, external electromagnetic fields, including coherent ones, thermal fluxes, and ionizing radiation.

A liquid crystal (LC) is a medium with additional internal degrees of freedom determined by the orientation of the long axes of interacting molecules. Orientational ordering can be changed by relatively weak external influences. This property leads to the representation of a liquid crystalline medium as soft or structured matter. The anisotropy of the physical properties of LCs leads to a significant complication of the mechanics and electrodynamics of anisotropic liquids and, consequently, to the emergence of many properties without analogues in ordinary isotropic substances.

Interest in the study of such media is determined by modern trends in the study of partially ordered media. In particular, analogies between the properties of LCs and their composites with biological objects are considered. Scientific and applied aspects include principles of controlling the motion of localized structures and inclusions in the LC environment, implemented using external fields or remotely, using coherent radiation.

Mechanisms of dynamic control of structures in LCs can be transferred to biological media with the inclusion of micro- and nanoparticles.

Methods and Algorithms: The main methods of investigation primarily include original techniques and approaches developed at the Institute of Laser Physics, Siberian Branch of the Russian Academy of Sciences, and other laboratories (LPI, Moscow; YSU, Yerevan), which, in our view, can be applied to the study of biological objects.

Among the well-known effects in LCs, it worth to note the Fredericks transition in constant magnetic and electric fields [1], more complex orientational dynamic effects [7], and their generalization in the form of light-induced reorientation [8]. However, the effects, where defects, special points where the orientation of the LC is not defined, are formed, may be more promising [4]. Sources can include nano-objects such as nanoparticles, polymer molecules [9], and even atomic structures from ionizing radiation [10]. Equally interesting are the processes of self-organization of nano- and microobjects due to the regular arrangement of defects [11] in periodic electric fields [4] or under the influence of laser radiation. In addition to mechanical inclusions creating defects, remote methods of generating and controlling LC structures are proposed using coherent sources [4, 12]. The mechanism for the appearance of such formations is the discovered thermo-orientational effect [13], caused by the reorientation of LC molecules by thermal and hydrodynamic flows. This report also considers a range of thermal and orientational structures that appear in absorbing LCs under the influence of localized radiation.

As analogies with processes occurring in biological systems, some examples from the hydrodynamics of liquid crystals can be given, which are quite complex phenomena due to the need to take into account the structural features of anisotropic fluid [14]. In particular, it is worth noting the Lehmann effect, leading to the rotation of molecules of cholesteric LC under the influence of a temperature gradient, as well as phenomena related to the flow of cholesteric under the action of weak pressure gradients, leading to "leakage" effects with a sharp increase in effective viscosity and the appearance of azimuthal velocity components.

For the interpretation of the effects presented in the perort, a theoretical model based on the representation of LC as a distributed medium is considered [15]. The main provisions of the proposed approach allow concluding the possibility of describing the liquid crystalline state of the medium in electromagnetic fields on nanoscale, including special orientation points and phase transitions in deformed structures [15, 16].

Results: The results of this work can serve as the listed properties of liquid crystalline media and their composites, considered from the point of view of some analogy with living objects, and the proposal for a deeper investigation of these analogies, such as biological liquid crystals.

Conclusion: The authors propose potential directions for interdisciplinary research.

1. Visualization of nano- and micro-objects in LCs, including biological molecules, as well as visualization of microstructures on the surface.
2. Determination of the nature of surface interaction of biological micro-objects upon contact with LC.
3. Investigation of the interaction between biological micro-objects.

Possible mechanisms:

a) Hydrodynamic impulse with the formation of a transmission channel.
b) Elastic orientation of signal without flow.
c) Electric impulse formed due to the flexoelectric effect.
d) Order parameter changes. Entropy (?) form.

e) Combinations of the previous mechanisms.

4. Control of biological micro-sized structures (formation, movement, rotation, etc.) based on techniques developed for LCs using electric fields, coherent radiation (laser tweezers), acoustics, thermal fluxes, and/or order parameter gradients.

5. Reconstruction of the object's shape during its visualization based on the orientational response of the LC. Determination of the degree of correctness for the inverse problem. Adaptation and refinement of LC theoretical models to describe partially ordered biological media in the presence of external influences.

In conclusion, it should be noted that effects in liquid crystalline structures or their analogies can be effectively utilized for research in the fields of biology, medicine, and other related sciences.

Funding: The research was supported by the Program of Fundamental Scientific Research in the Russian Federation for a long-term period (2021–2030), project No. 1021062211014-7-1.3.6.

Список литературы/References

1. Kleman M., Lavrentovich O. Soft Matter Physics. Springer, 2003
2. Lowe A.M., Abbott N.L. Liquid Crystalline Materials for Biological Applications. *Chem Mater*. 2012;24(5):746-758
3. Woltman S.J., Jay G.D., Crawford G.P. Liquid-crystal materials find a new order in biomedical applications. *Nat Mater*. 2007;6:929-938
4. Bagayev S.N., Klementyev V.M., Nyushkov B.N., Pivtsov V.S., Trashkeev S.I. New methods of highly efficient controlled generation of radiation by liquid crystal nanostructures in a wide spectral range. *J Phys Conf Ser*. 2012;345:012018. doi 10.1088/1742-6596/345/1/012018
5. Doostmohanmadi A., Ladoux B. Physics of liquid crystals in cell biology. *ArXiv*. 2021
6. Rey A.D. Liquid crystal models of biological materials and processes. *Soft Matter*. 2010;6:3402-3429
7. Trahkeev S.I., Klementev V.M., Statsenko P.A. Determinate and stochastic orientation dynamics of nematic liquid crystals in the cross electric fields with requencies up to the megahertz range. *Liq Cryst*. 2006;33(4):417-438
8. Kitaeva V.F., Zolot'ko A.S., Sobolev N.N. Self-focusing of laser radiation in the case of a Fredericks transition. *Sov Phy. Usp*. 1982;25(10):758-760. doi 10.1070/PU1982v025n10ABEH004611
9. Trashkeev S.I., Grachev G.N., Pozdnyakov G.A. Nanostructures in a nematic liquid crystal. *Tech Phys*. 2007;52(9):1188-1194
10. Trashkeev S.I., Grachev G.N., Pozdnyakov G.A. Observation of the Electrostatic Defects Induced by Ionizing Radiation in a Liquid Crystal. *Tech Phys Lett*. 2007;33(5):438-440
11. Klementev V.M., Pozdnyakov G.A., Trashkeev S.I. Methods for visualization and ordering of nanoobjects in a liquid crystalline medium. *Zavodskaya Laboratoriya. Diagnostika Materialov*. 2008;74(3):32-38 (in Russian)
12. Shvetsov S.A., Zolot'ko A.S., Voronin G.A. et al. Light-induced umbilical defects due to temperature gradients in nematic liquid crystal with a free surface. *Opt Mater Express*. 2021;11(6):1705-1712. doi 10.1364/OME.425926
13. Trashkeev S.I., Britvin A.V. Thermal Orientation Effect in a Nematic Liquid Crystal. *Tech Phys*. 2011;56(6):747-753
14. Chandrasekhar S. Liquid Crystals. Cambridge University Press, 1993
15. Kudryavtsev A.N., Purtov P.A., Trashkeev S.I. Nonlocal tensor order parameter of the deformed state of liquid crystals. *arXiv*. 2017
16. Trashkeev S.I., Statsenko P.A., Khomyakov M.N., Shvetsov S.A. Non-local Q-tensor approach for description of elastic deformations of nematic liquid crystals at sub-micronscale. *Liq Cryst Their Appl*. 2023;23(3):66-76 (in Russian)

doi 10.18699/bgrs2024-9.6-06

О патогенезе отека стромы щитовидной железы

Ушаков А.[1]*, Вельмякина С.[1], Небрат В.[2]

[1] *Клиника доктора А.В. Ушакова, Москва, Россия*

[2] *Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал ИЦиГ СО РАН, Новосибирск, Россия*

* *docthyroid@gmail.com*

Ключевые слова: щитовидная железа; ультразвуковое исследование; отек стромы; интерстиций; автономная нервная система; ионная система интерстиция

Мотивация и цель: Ультразвуковое исследование (УЗИ) позволяет по эхогенности оценивать величину насыщенности жидкостью паренхимы ЩЖ. В том числе, выявлять три варианта накопления жидкости в интерстиции: распространенный, сегментарный и линейный [1]. Это явление выглядит при УЗИ темно-серым (гипоэхогенным), может иметь разную выраженность, увеличиваться или уменьшаться до нормы (изоэхогенным). Предыдущие исследования показали отсутствие абсолютной прямой связи тироидной гипоэхогенности с гормональным обменом, аутоиммунным процессом и интенсивностью кровотока в железе [1, 2]. Вместе с тем, одинаковая гипоэхогенность ЩЖ встречается при первичном гипотиреозе и диффузном гипертиреозе (болезни Грейвса) [1]. Такое противоречие, с одной стороны, уменьшает диагностическую ценность УЗИ, а с другой стороны, акцентирует внимание на патогенез тиреоидной гипоэхогенности. Целью этой работы стало выявление патогенетических принципов и особенностей отека стромы ЩЖ при УЗИ.

Методы и алгоритмы: Изучение патогенеза отека стромы ЩЖ выполнялось на основе клинического материала – 3040 пациентов за 16 лет в специализированной тиреоидологической клинике, где для оценки гипоэхогенности ткани ЩЖ использовались данные УЗИ-исследования совместно с анализами крови, анамнезом болезни, и текущим статусом пациентов. В том числе, у 1668 пациентов контролировали ультразвуком состояние ЩЖ после нескольких курсов лазеротерапии и фототерапии. Для исследования особенностей патогенеза отека стромы ЩЖ проводился поиск и анализ научной литературы в PubMed, Google Scholar, и фондах научно-медицинских библиотек по тегам «отек стромы ЩЖ», «гипоэхогенность ЩЖ», «УЗИ диффузных изменений ЩЖ», «электроосмос», «отек эпителиальных тканей» и другие.

Результаты и обсуждение: Накопление жидкости в интерстиции ЩЖ принято традиционно называть отеком стромы. Более известным вариантом такого отека является острый отек около магистральных сосудов (обычно 3 или 4 порядка), который возникает в ближайший час после пункции железы, внутривенного тромболизиса или пункции подключичной артерии [3, 4]. Этот вариант отека стромы (линейный) определяется при УЗИ черными полосами, может сопровождаться увеличением доли в 1.5–3 раза и местной болью. Такой отек быстро уменьшается и исчезает в течение 1–2 суток. Другим вариантом отека стромы ЩЖ является накопление жидкости в интерстиции всей доли или отдельно в ее сегментах (крупных, средних или малых (дольках)), и эти явления

сохраняются продолжительно (месяцами) [1]. В исследованиях показано, что при физиотерапии с помощью инфракрасного лазера или фототерапии (в том числе гелиотерапии) выраженность распространенной и сегментарной гипоэхогенности может уменьшаться до изоэхогенности в течение нескольких недель [1, 5, 6].

Если химический состав внутрисосудистой и интерстициальной жидкости во всех сегментах ЩЖ одинаков, то основа патогенеза отека стромы может быть связана с физическими процессами в железе. Исследования последних лет подтверждают, что механизм острого и хронического отека интерстиция (стромы) ЩЖ связан с физическим процессом – электросмосом [7]. Острый линейный и хронический распространенный отек ЩЖ, связывается с околососудистым проникновением жидкости [8]. Линейный – вдоль относительно крупных сосудов (2–4 порядка), создавая при УЗИ картину линейной гипоэхогенности. Хронический распространенный отек охватывает мелкие сосуды, демонстрируя общую гипоэхогенность долей ЩЖ. Соответственно, возникают следующие ключевые вопросы. Почему острый паравазальный отек возникает и исчезает после пункции доли железы (в редких случаях)? Как фототерапия способствует устранению хронического отека стромы ЩЖ? Какая основа механизма отека сегментов ЩЖ? Для ответа необходимо обратить внимание на два важных условия функционирования любых органов, включая ЩЖ. Первое: сегменты органов имеют относительно обособленную иннервацию и васкуляризацию, нервные волокна всегда сопровождают сосуды и нервный импульс имеет электрическую природу. Второе – интерстициальные каналы во всем организме и его органах представляют систему циркуляции ионов воды (H^+ и OH^-), функционально взаимосвязанную с нервно-сосудистой сетью [9]. Эти два условия относятся ко всему организму. С этой точки зрения можно предположить, что регуляция процессов в интерстиции организована в соответствии с сегментарной структурой организма. Двумя сопряженными регуляторными системами такой регуляции, как видно, является автономная нервная система (АНС) и ионная система интерстиция (ИСИ). Можно думать, что изменение потоков ионов и их отношений в системе межорганных, паравазальных и интерстициальных каналов шеи оказывает влияние на величину возбуждения регионарной АНС, сегментарно регулирующую ткань и сосудистую (особенно артериальную) сеть ЩЖ, что и приводит к общему или локальному отеку стромы железы. Предположительно, ИСИ имеет ведущее значение в таком процессе. Вместе с тем, АНС может иметь ключевое значение в развитии сегментарных и паравазальных отеков за счет непосредственного электрического влияния на локальный ионный баланс в тиреоидной паренхиме. Возможно, острый отек стромы появляется и вскоре исчезает в связи с кратковременным раздражением готовых к ответной реакции местных АНС и ИСИ.

Хронический отек стромы ЩЖ, как видно, представляет более устойчивые изменения АНС и ИСИ. Особенности вовлечения разных сегментов в отек стромы, как видно, зависят от сегментарного распределения в железе АНС и ИСИ. Влияние фототерапии или лазеротерапии на отек стромы, предположительно, в основном связан с ИСИ [10]. Отсутствие отека стромы ЩЖ у пятой части людей с относительно малым избытком ТТГ, АТ-ТПО и АТ-ТГ [2], вероятно, связано с нормальной циркуляцией ионов в интерстиции. Прямая зависимость большего избытка ТТГ, АТ-ТПО и АТ-ТГ почти в 60 % случаев [2] может быть связана с большим усилением напряжения ткани ЩЖ под влиянием АНС и с участием

ИСИ. Нормальный уровень ТТГ, АТ-ТПО и АТ-ТГ еще у 20 % пациентов с гипоэхогенной ЩЖ [2], предположительно, представляет временный период после продолжительного гиперфункционирования (перенапряжения) ЩЖ, в течение которого уменьшена нервная стимуляция и клетки органа функционально истощены, но сохраняются изменения ИСИ. При этом возможности УЗИ-контроля дают преимущество в обнаружении признаков отека стромы ЩЖ; в мониторинге процесса лечения, в определении эффективности способов и методик лечения, влияющих на АНС и ИСИ. 16-летний период ультразвуковой и лечебной (лазеро- и фототерапевтической) специализированной практики и данные литературы позволяют быть уверенными в предложенном механизме тиреоидного отека стромы. Незначительная сложность сводится к дифференциальной диагностике специалистами ультразвуковой гипоэхогенности отека стромы от лимфоцитарной инфильтрации [1].
Выводы: Патогенез отека стромы ЩЖ, вероятно, происходит электроосмотически под влиянием АНС и ИСИ (ионной системы интерстиция). УЗИ позволяет контролировать выраженность отека интерстиция стромы ЩЖ.
Благодарность: Авторы благодарны сотрудникам клиники, участвовавшим в приеме пациентов: Татьяне Денисовой и Екатерине Анисковой.

Pathogenesis of thyroid stromal swelling

Ushakov A.[1]*, Velmyakina S.[1], Nebrat V.[2]
[1] *Ushakov Thyroid Clinic, Moscow, Russia*
[2] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
* *docthyroid@gmail.com*

Key words: thyroid gland; ultrasound; stromal swelling; interstitium; autonomic nervous system; interstitial ion system

Motivation and Aim: Ultrasound examination (ultrasound) allows assessing the amount of fluid saturation of the thyroid parenchyma by echogenicity. In particular, to identify three variants of fluid accumulation in the interstitium: widespread, segmental and linear [1]. This phenomenon appears dark gray (hypoechoic) on ultrasound, may have different severity, increase or decrease to normal (isoechoic). Previous studies have shown the absence of an absolute direct connection between thyroid hypoechogenicity and hormonal metabolism, the autoimmune process and the intensity of blood flow in the gland [1, 2]. At the same time, the same hypoechogenicity of the thyroid gland occurs in primary hypothyroidism and diffuse hyperthyroidism (Graves disease) [1]. This contradiction, on the one hand, reduces the diagnostic value of ultrasound, and on the other hand, focuses attention on the pathogenesis of thyroid hypoechogenicity. The purpose of this work was to identify the pathogenetic principles and features of swelling of the thyroid stroma during ultrasound.
Methods and Algorithms: The study of the pathogenesis of thyroid stromal swelling was carried out on the basis of clinical material – 3040 patients over 16 years in a specialized thyroid clinic, where ultrasound data were used in conjunction with blood tests, medical history, and the current status of the patients to assess the hypoechogenicity of the thyroid tissue. In particular, in 1668 patients the thyroid condition was monitored by

ultrasound after several courses of laser therapy and phototherapy. To study the features of the pathogenesis of thyroid stromal swelling, a search and analysis of scientific literature was carried out in PubMed, Google Scholar, and the collections of scientific and medical libraries using the tags “thyroid stromal swelling”, “hypoechogenicity of the thyroid gland”, “ultrasound of diffuse changes in the thyroid gland”, “electroosmosis”, “swelling of epithelial tissues” and others.

Results and Discussion: The accumulation of fluid in the thyroid interstitium is traditionally called stromal swelling. A more well-known variant of such edema is acute swelling near the great vessels (usually 3 or 4 orders), which occurs within the next hour after gland puncture, intravenous thrombolysis or puncture of the subclavian artery [3, 4]. This variant of stromal swelling (linear) is determined by ultrasound as black stripes and can be accompanied by an increase in the lobe by 1.5–3 times and local pain. This swelling quickly decreases and disappears within 1–2 days. Another option for thyroid stroma swelling is the accumulation of fluid in the interstitium of the entire lobe or separately in its segments (large, medium or small (lobules)), and these phenomena persist for a long time (months) [1]. Studies have shown that with physical therapy using infrared laser or phototherapy (including heliotherapy), the severity of widespread and segmental hypoechogenicity can decrease to isoechogenicity within a few weeks [1, 5, 6].

If the chemical composition of intravascular and interstitial fluid in all segments of the thyroid gland is the same, then the basis of the pathogenesis of stromal swelling may be associated with physical processes in the gland. Recent studies confirm that the mechanism of acute and chronic swelling of the thyroid interstitium (stroma) is associated with a physical process – electrosmosis [7]. Acute linear and chronic widespread thyroid swelling is associated with perivascular fluid penetration [8]. Linear – along relatively large vessels (2–4 orders), creating a picture of linear hypoechogenicity on ultrasound. Chronic widespread swelling covers small vessels, demonstrating the general hypoechogenicity of the thyroid lobes. Accordingly, the following key questions arise. Why does acute paravasal swelling appear and disappear after puncture of the gland lobe (in rare cases)? How does phototherapy help eliminate chronic swelling of the thyroid stroma? What is the basis of the mechanism of swelling of thyroid segments?

To answer, you need to pay attention to two important conditions for the functioning of any organs, including the thyroid gland. First: organ segments have relatively separate innervation and vascularization, nerve fibers always accompany the vessels and the nerve impulse is electrical in nature. Secondly, interstitial channels throughout the body and its organs represent a system of circulation of water ions (H+ and OH-), functionally interconnected with the neurovascular network [9]. These two conditions apply to the entire body. From this point of view, it can be assumed that the regulation of processes in the interstitium is organized in accordance with the segmental structure of the body. Two associated regulatory systems for such regulation, as can be seen, are the autonomic nervous system (ANS) and the interstitial ion system (ISI). One might think that changes in ion flows and their relationships in the system of interorgan, paravasal and interstitial channels of the neck influence the magnitude of excitation of the regional ANS, segmentally regulating tissue and the vascular (especially arterial) network of the thyroid gland, which leads to general or local swelling of the gland stroma. Presumably, ISI plays a leading role in such a process. At the same time, the ANS may be of key importance in the development of segmental and paravasal swelling due to the direct

electrical effect on the local ionic balance in the thyroid parenchyma. Perhaps acute stromal swelling appears and soon disappears due to short-term irritation of the local ANS and ISI, ready to respond.

Chronic swelling of the thyroid stroma appears to represent more persistent changes in the ANS and ISI. The features of the involvement of different segments in stromal swelling, as can be seen, depend on the segmental distribution of the ANS and ISI in the gland. The effect of phototherapy or laser therapy on stromal swelling is believed to be mainly related to ISI [10]. The absence of thyroid stromal swelling in a fifth of people with a relatively small excess of thyroid-stimulating hormone (TSH), antibodies to thyroid peroxidase and thyroglobulin (TPOAb, TGAb) [2] is probably due to the normal circulation of ions in the interstitium. The direct dependence of a greater excess of TSH, TPOAb and TGAb in almost 60 % of cases [2] can be associated with a greater increase in thyroid tissue tension under the influence of the ANS and with the participation of ISI. Normal levels of TSH, TPOAb and TGAb in another 20 % of patients with a hypoechoic thyroid gland [2] presumably represent a temporary period after prolonged hyperfunctioning (overstrain) of the thyroid gland, during which nerve stimulation is reduced and the cells of the organ are functionally depleted, but remain intact. ISI changes. At the same time, the capabilities of ultrasound control provide an advantage in detecting signs of swelling of the thyroid stroma; in monitoring the treatment process, in determining the effectiveness of treatment methods and techniques that affect the ANS and ISI. A 16-year period of ultrasound and therapeutic (laser and phototherapeutic) specialized practice and literature data allow us to be confident in the proposed mechanism of thyroid stromal swelling. A minor difficulty comes down to the differential diagnosis by ultrasound specialists of hypoechogenicity of stromal swelling from lymphocytic infiltration [1].

Conclusion: The pathogenesis of thyroid stromal swelling probably occurs electroosmotically under the influence of the ANS and ISI. Ultrasound allows you to monitor the severity of swelling of the interstitium of the thyroid stroma.

Acknowledgement: The authors are grateful to the clinic staff who participated in the reception of patients: Tatyana Denisova and Ekaterina Aniskova.

Список литературы/References

1. Ushakov A.V. Principles and features of ultrasound hypoechogenicity in diffuse thyroid pathology. *Quant Imaging Med Surg.* 2024;14(3):2655-2670
2. Trimboli P. et al. One in five subjects with normal thyroid ultrasonography has altered thyroid tests. *Endocr J.* 2012;59(2):137-143
3. Shi L., Yan C., Xu W., Huang P. Acute diffuse and transient thyroid swelling after intravenous thrombolysis for acute ischemic stroke: A case report. *Medicine (Baltimore).* 2018;97(36):e12149
4. Zhu T., Yang Y., Ju H., Huang Y. Acute thyroid swelling after fine needle aspiration-a case report of a rare complication and a systematic review. *BMC Surg.* 2021;21(1):175
5. Ushakov A. Case of Graves' disease recovery. *J Clin Transl Endocrinol Case Rep.* 2023;27:100139
6. Höfling D.B. et al. Low-level laser therapy in chronic autoimmune thyroiditis: a pilot study. *Lasers Surg Med.* 2010;42(6):589-596
7. Rame J.E., Müller J. Myocardial Edema Revisited in a New Paradigm of Cardiac Electrical Microcurrent Application in Heart Failure. *Bioelectricity.* 2021;3(3):171-175
8. Li H. et al. Layers of interstitial fluid flow along a "slit-shaped" vascular adventitia. *J Zhejiang Univ Sci B.* 2021;22(8):647-663
9. Nebrat V. Human water model: interstitium and meridians of traditional Chinese medicine. In: Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022). Novosibirsk, 2022;816. doi 10.18699/SBB-2022-474
10. Cha Y., Choe H., Oh S. Treatment of patient with cirrhosis-ascites by the mineral pulse light stimuli on LV acupoints-a case report. *Res Square.* 2020. doi 10.21203/rs.3.rs-52673/v1

doi 10.18699/bgrs2024-9.6-07

General principles of structure and ultrastructural differences in the interstitial pathways of tissue fluid movement

Bgatova N.[1]*, Skudin N.[2], Serafimov S.[3], Chernykh D.[4], Letyagin A.[1]

[1] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia*

[2] *Novosibirsk Regional Clinical Oncology Center, Novosibirsk, Russia*

[3] *Novosibirsk Medical University, Novosibirsk, Russia*

[4] *S.N. Fyodorov Federal State Institution National Medical Research Center Intersectoral Research and Technology Complex "Eye Microsurgery" Ministry of Health of the Russian Federation, Moscow, Russia*

* *nataliya.bgatova@yandex.ru*

Key words: colon; gums; eye; synovial membrane; interstitium; ultrastructure

Motivation and Aim: Not being part of the lymphatic system, the interstitium is an important link in the continuous circulation of tissue fluid (carrying metabolic substrates, waste products and immune mediators) and it is the initial link in the lymphatic drainage of the internal environment of the body [1]. The structure and composition of the interstitium in different organs and tissues have common structural features and can also vary. The purpose of the study was to identify the features of the structural organization of the interstitium of various organs under normal and pathological conditions.

Methods and Algorithms: The study used archival materials carried out in different years to study the structure of the mucous membrane of the gum, adenocarcinoma of the colon, eye (cornea, ciliary body, choroid), human synovial membrane. Organ fragments were fixed in a 4 % solution of paraformaldehyde prepared in Hanks' medium, further fixed for 1 hour in a 1 % solution of osmium tetroxide (OsO4) in phosphate buffer pH 7.4, dehydrated in ethyl alcohol of increasing concentrations and embedded in Epon. Ultrathin sections with a thickness of 70–100 nm were obtained on a Leica EM UC7 ultratome (Leica Microsystems, Germany) and contrasted with a saturated aqueous solution of uranyl acetate and lead citrate. Digital photographs were taken using a JEM 1400 electron microscope (JEOL, Japan) at the Center for Microscopic Analysis of Biological Objects of the Siberian Branch of the Russian Academy of Sciences. Morphometric analysis was carried out using the Image J computer program (Wayne Rasband, USA). The mean (M) and standard deviation (SD) were calculated using Microsoft Excel software (Microsoft, USA). The significance of differences between the studied parameters was determined using Statistica 6.0 software (StatSoft, USA) and the Mann-Whitney U test at a confidence level of 95 % ($P < 0.05$).

Results: The interstitium makes up about 20 % of body weight. Interstitial spaces exist between the cells of any organ. The structure of the interstitium is usually represented by a framework of three-dimensional fibrous collagen network, in which collagen fibers form thick bundles [2]. The interstitium is considered a structure of the extracellular matrix in which the main substance is formed from proteoglycans, consisting of glycosaminoglycans and free hyaluronan. They form gel-like structures due to their ability to attract water due to their high polarity. Glycosoaminoglycans interact with the collagen structure of the extracellular matrix but can also be found in the interstitial fluid. Interstitial fluid is an ultrafiltrate of plasma and containing protein. The structure of non-vascular pathways for the movement of interstitial fluid may have similarities and vary

in different organs. There are interstitial fluid of tissue channels, fluid of hydrophobic channels, fluid of hydrophilic channels, perivascular fluid, cerebrospinal fluid and interstitial fluid of fascia [3]. When analyzing the interstitium of the studied organs, differences in its structural organization were noted (Fig. 1).

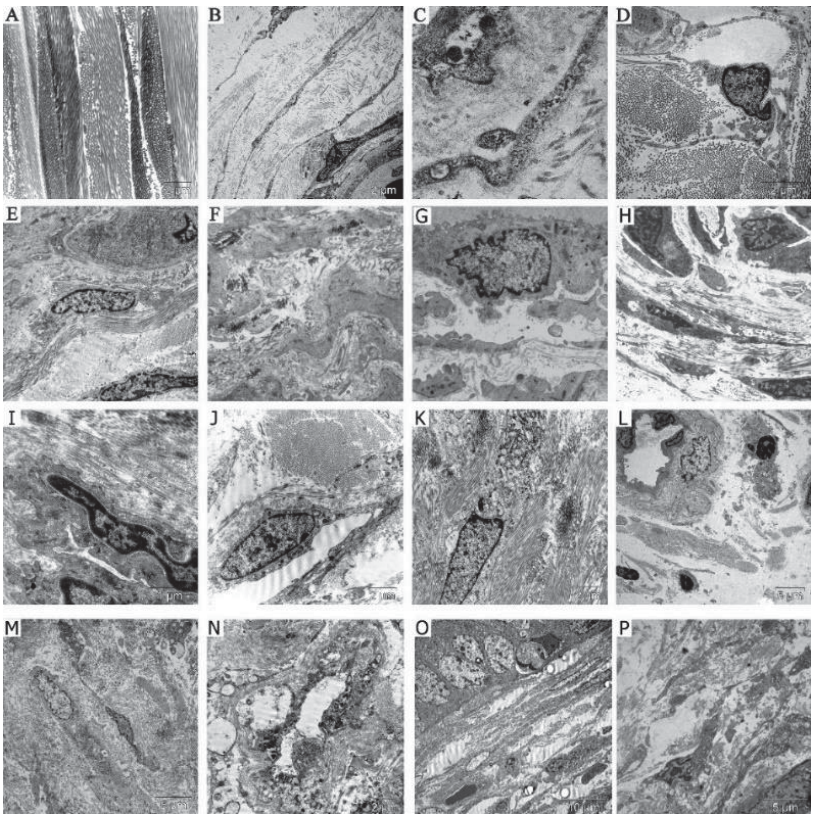


Fig. 1. Various states of interstitial pathways of tissue fluid movement in the structures of the eye, synovium, gums and colon. A – narrow spaces between parallel oriented collagen fibers in the cornea. B – wide interstitial spaces in the ciliary body formed by unidirectional collagen fibers and limited by fibroblasts and their processes. C, D – channels formed by fibroblast-like cells in the choroid (C) and in the ciliary body (D). E – interstitial spaces in the synovium formed by parallel oriented collagen fibers. F – multidirectional arrangement of collagen fibers in the interstitial space of the synovial membrane in stage I arthrosis. G, H – wide interstitial spaces limited by collagen fibers and fibroblast-like cells in synovial membrane in stage II arthrosis. I – narrow pericapillary and interstitial spaces in the gingival mucosa. J – expanded pericapillary and interstitial spaces in the gingival mucosa during inflammation. K – multidirectional arrangement of collagen fibers in the gum interstitium during inflammation. L – wide interstitial spaces limited by collagen fibers in the gingival mucosa during inflammation. M – moderate density and unidirectionality of collagen fibers in the interstitium of moderately differentiated colon adenocarcinoma. N – multidirectional arrangement of collagen fibers in the interstitium of well-differentiated colon adenocarcinoma. O – formation of narrow channels in the interstitium of moderately differentiated colon adenocarcinoma. P – gluing of collagen fibers and formation of wide channels in the extracellular matrix of poorly differentiated colon adeno-carcinoma

The common features of the studied organs interstitium structure under normal conditions were the presence between cells and vessels of narrow spaces formed by parallel collagen fibrils, which determine the direction of movement of tissue fluid. A feature of the choroid and synovium was the presence in the interstitium of channels formed by collagen fibers and fibroblast-like cells. Similar channels have been described in the structure of the interstitium of the extra-hepatic bile ducts [4]. Under pathological conditions, an increase in pericapillary spaces, disorientation of collagen fibers, and expansion of interstitial spaces were observed. Analysis of co-lon adenocarcinoma revealed abnormal remodeling of interstitial collagen, resulting in excessive collagen deposition, fibril adhesions, and the wide interstitial channels formation (Fig. 1).

Conclusion: When analyzing the interstitium of the studied organs, differences in its structural organization were noted. Under normal conditions, narrow pericapillary and interstitial spaces formed by parallels located unidirectional collagen fibrils were noted. Under pathological conditions, disorientation of collagen fibers and expansion of interstitial spaces were observed. In the choroid and synovium, the presence of interstitial channels bounded by collagen fibers and fibroblast-like cells was detected. Under conditions of malignant growth, there was excessive deposition of collagen, gluing of collagen fibers with the formation of channels for migration of tumor cells.

Funding: The study was supported by budgetary funding from the Russian Government for basic scientific research of the Research Institute of Clinical and Experimental Lymphology as part of a state order NFWNR-2022-0012.

References

1. Borodin Iu.I. Regional lymphatic drainage and lymphodetoxication. *Morfologiia.* 2005;128(4):25-28 (in Russian)
2. Dargent A., Dumargne H., Labruyère M. et al. Role of the interstitium during septic shock: a key to the understanding of fluid dynamics? *J Intensive Care*. 2023;11(1):44. doi 10.1186/s40560-023-00694-z
3. Liu W.T., Cao Y.P., Zhou X.H., Han D. Interstitial Fluid Behavior and Diseases. *Adv Sci (Weinh*). 2022;9(6):e2100617. doi 10.1002/advs.202100617
4. Benias P.C., Wells R.G., Sackey-Aboagye B. et al. Structure and Distribution of an Unrecognized Interstitium in Human Tissues. *Sci Rep.* 2018;8(1):4947. doi 10.1038/s41598-018-23062-6

doi 10.18699/bgrs2024-9.6-08

Features of extracellular matrix remodeling in human uveal melanoma

Taskaeva Iu.S.[1]*, Shatruk A.Yu.[1], Bgatova N.P.[1], Yeremina A.V.[2], Trunov A.N.[2], Chernykh V.V.[2]

[1] *Laboratory of Ultrastructural Research, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *S.N. Fyodorov Federal State Institution National Medical Research Center Intersectoral Research and Technology Complex "Eye Microsurgery" Ministry of Health of the Russian Federation, Novosibirsk, Russia*

** inabrite@yandex.ru*

Key words: uveal melanoma; extracellular matrix; cancer-associated fibroblasts; matrix metalloproteina-ses

Motivation and Aim: The tumor microenvironment (TME) is a specific heterogeneous environment surrounding the tumor, consisting of blood and lymphatic vessels, stromal and immune cells, progenitor cells, signaling molecules and extracellular matrix (ECM) [1]. The interaction of tumor cells with TME components promotes the activation of signaling pathways aimed at carcinogenesis, evasion of immune surveillance and the development of drug resistance [2].

A key element of the TME is the ECM, an organized three-dimensional network of macromolecules that maintains the integrity of tissues and organs. The architecture of the ECM includes the interstitial matrix and basement membrane. The interstitial matrix mainly consists of fibrillar collagens, fibronectin, elastin, hyaluronic acid and other components [3]. The main source of synthesis of ECM components are activated cancer-associated fibroblasts (CAFs) [4]. The ECM contributes to the formation of a mechanical framework for tumor cells, participates in the regulation of signaling cascades that influence the phenotype of stromal and immune cells, and modulates the processes of metastasis, angiogenesis and proliferation of tumor cells [4].

ECM remodeling is a complex process of degradation and renewal of a network of macromolecules [5], however, the factors secreted by tumor cells and matrix metalloproteinases (MMPs) produced by CAFs can affect the ECM to support the processes of carcinogenesis and the development of drug resistance [5].

Uveal melanoma (UM) is the most common intraocular malignancy arising from choroidal pigment cells. It is known that UM metastases are diagnosed in 50 % of patients, regardless of the type of treatment [6] and are the main cause of death in patients with UM. Despite advances in the development of new therapies for UM, the average survival rate for metastatic disease is approximately 5–10 months [6].

It has been shown that the TME, in particular the ECM, regulates signaling pathways associated with metastasis and the development of resistance to therapy [7]. It is assumed that tumors are capable of forming a unique tumor-specific architectural type of ECM: some types of cancer are characterized by the accumulation of ECM components to regulate the signaling pathways of angiogenesis, growth and proliferation; in other cases, the processes of degradation of ECM components are increased to facilitate invasion and metastasis [7]. For example, UM is characterized by the destruction of collagen types I

and IV under the influence of MMP2, MMP9 and MMP14, secreted by CAFs [8]. Despite advances in studies of UM carcinogenesis, currently the role of individual ECM components in the pathogenesis of UM is not well understood and requires detailed study [9]. Thus, the study of the structural features and processes of reorganization of the ECM can contribute to the development of new approaches to the treatment of UM. The aim of this study was to evaluate the expression of markers associated with ECM remodeling in UM.

Methods and Algorithms: The archived materials were received from 15 patients with uveal melanoma who had undergone curative enucleation in S. Fyodorov Eye Microsurgery Federal State Institution, Novosibirsk, Russia. The study was approved by the local ethics committee of the RICEL – branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (Protocol No. 180 of 04/28/2023).

The formalin-fixed, paraffin-embedded sections from post-equatorial zone of the choroid and intratumoral area of the tumor samples were dewaxed and rehydrated and submitted to heat-induced epitope retrieval (citrate buffer, pH 6.0). Sections were stained using anti-αSMA (Abcam, UK), Collagen I (Invitrogen, USA), FN1 (Elabscience, China), MMP2 (Abcam, UK), MMP14 (Elabscience, China), TIMP2 (Abcam, UK) primary antibodies. The images were captured with an Axio Scope microscope A1 (ZEISS, Germany) using an Axiocam 512 color CCD camera (ZEISS, Germany) and ZEN 2.3 software (ZEISS, Germany) at 40× magnification. Microscopic analysis was carried out at the Multiple-access Center for Microscopy of Biological Subjects (Institute of Cytology and Genetics, Novosibirsk, Russia).

The analysis of digital images of the choroid and tumor (10 fields per patient) was performed using Image J software (Wayne Rasband, USA). Stained choroid and tumor sections were assessed using an immunoreactive scale. Scores were established for the intensity and percentage of positive staining of the cytoplasm. The extent of staining (area) was scored as "0" (< 1 %), "1" (1–25 %), "2" (>25–50 %), "3" (> 50–75 %), or "4" (> 75 %). The staining intensity was scored as "0" (no staining), "1" (weak staining), "2" (moderate staining), or "3" (strongly staining). The weighted scores were calculated by the formula: area × intensity of staining.

The significance of differences between the studied parameters was determined using the Statistica 6.0 software (StatSoft, USA) and the Mann-Whitney U-test (non-parametric statistics) at a 95 % significance level ($p < 0.05$). Data are presented as mean and standard deviation (M±SD).

Results: An analysis of the expression of proteins associated with ECM remodeling and an assessment of the area of $\alpha SMA^{(+)}$ fibroblasts were performed. The area of $\alpha SMA^{(+)}$ staining zones indicating CAFs activity in the tumor was 4 times higher than in the choroid ($p < 0.0005$). Type I collagen expression levels were 76.5 % higher ($p < 0.005$) in the post-equatorial choroid compared to the UM. When assessing the expression of FN (FN1), it was revealed that the expression level of this protein was 4.6 times higher in UM samples ($p < 0.00005$) compared to the choroid. Tumor samples were also characterized by high expression of MMPs and tissue inhibitor of MMPs, compared with the choroid: MMP2 expression by 57 % ($p < 0.05$), MMP14 by 62 % ($p < 0.005$), and TIMP2 2 times ($p < 0.05$).

Conclusion: It is known, that αSMA (alpha smooth muscle actin) represents a key marker of CAFs activation [4]. Furthermore, αSMA can be expressed on the surface of vascular smooth muscle cells; however, in UM, the proliferation of the vascular network occurs

due to the formation of small blood vessels, which are characterized by the absence of smooth muscle cells [6]. In this study, when analyzing postequatorial choroidal samples, αSMA$^{(+)}$ staining was observed predominantly along large blood vessels, while stromal cells were stained in the intratumoral zone of the tumor. Thus, it can be assumed that a high content of αSMA$^{(+)}$ zones in the tumor may indicate high functional activity of CAFs. ECM remodeling occurs primarily through the regulation of the production of MMPs by CAFs. Tissue inhibitors of matrix metalloproteinases (TIMPs) regulate the activity of MMPs by forming non-covalent complexes, thereby limiting tissue destruction. It has been shown that the formation of the MMP14-MMP2-TIMP2 complex activates MMP2 and stimulates the degradation of collagen fibers and ECM remodeling to enhance the processes of angiogenesis and tumor metastasis [10]. The studied tumor samples were characterized by high levels of expression of proteins associated with ECM degradation – fibronectin, MMP14, MMP2 and TIMP2. Thus, the results obtained may indicate an active process of ECM remodeling, which may affect the invasive ability and metastasis of tumor cells and requires further study.

Funding: This work was supported with finances from the Research Institute of Clinical and Experimental Lymphology (Novosibirsk) as part of state order, No. NFWNR-2022-0012.

References

1. Walker C., Mojares E., Del Río Hernández A. Role of extracellular matrix in development and cancer progression. *Int J Mol Sci*. 2018;19(10):3028. doi 10.3390/ijms19103028
2. Hanker A.B., Estrada M.V., Bianchini G. et al. Extracellular matrix/integrin signaling promotes resistance to combined inhibition of HER2 and PI3K in HER2+ breast cancer. *Cancer Res.* 2017;77(12):3280-3292. doi 10.1158/0008-5472.CAN-16-2808
3. Manou D., Caon I., Bouris P. et al. The complex interplay between extracellular matrix and cells in tissues. *Methods Mol Biol*. 2019;1952:1-20. doi 10.1007/978-1-4939-9133-41
4. Biffi G., Tuveson D.A. Diversity and biology of cancer-associated fibroblasts. *Physiol Rev*. 2021;101(1):147-176. doi 10.1152/physrev.00048.2019
5. Dzobo K., Dandara C. The extracellular matrix: its composition, function, remodeling, and role in tumorigenesis. *Biomimetics (Basel)*. 2023;8(2):146. doi 10.3390/biomimetics8020146
6. Kaliki S., Shields C.L., Shields J.A. Uveal melanoma: estimating prognosis. *Indian J Ophthalmol.* 2015;63(2):93-102. doi 10.4103/0301-4738.154367
7. Cox T.R. The matrix in cancer. *Nat Rev Cancer*. 2021;21(4):217-238. doi 10.1038/s41568-020-00329-7
8. Lai K., Conway R.M., Crouch R., Jager M.J., Madigan M.C. Expression and distribution of MMPs and TIMPs in human uveal melanoma. *Exp Eye Res.* 2008;86(6):936-941. doi 10.1016/j.exer.2008.03.010
9. Goesmann L., Refaian N., Bosch J.J., Heindl L.M. Characterization and quantitation of the tumor microenvironment of uveal melanoma. *Biology (Basel)*. 2023;12(5):738. doi 10.3390/biology12050738
10. Duan F., Peng Z., Yin J., Yang Z., Shang J. Expression of MMP-14 and prognosis in digestive system carcinoma: a meta-analysis and databases validation. *J Cancer*. 2020;11(5):1141-1150. doi 10.7150/jca.36469

10

Симпозиум «Общие проблемы в изучении когнитивных процессов; модели когнитивной деятельности»

Symposium “General problems in the study of cognitive processes; models of cognitive activity”

1847

doi 10.18699/bgrs2024-10-01

Диагностика когнитивных искажений в публичных, групповых и персональных текстовых коммуникациях

Ариничева А.[1]*, Колонин А.[1, 2]

[1] *Новосибирский государственный университет, Новосибирск, Россия*

[2] *Aigents, Новосибирск, Россия*

* *a.arinicheva@g.nsu.ru*

Ключевые слова: когнитивные искажения; поиск оптимальных параметров; анализ данных

Мотивация и цель: Улучшение качества жизни среднестатистического человека, нестабильная ситуация в мире и широкая доступность интернет-ресурсов привели в первую очередь к резкому увеличению количества информации, которую потребляет человек в течение дня, включая как проверенные факты, так и ложные утверждения, в числе которых манипулятивные тексты, направленные на умышленное искажение картины мира и психического состояния их потребителя. Виртуальное пространство усугубляет трудности отличия достоверной информации от фальсификации из-за отсутствия возможности наблюдения за мимикой, жестами и интонацией собеседника, что характерно для личного общения. В такой ситуации человеку, столкнувшемуся с вызывающей сомнения новостью, способен помочь ассистент, который проанализирует содержание статьи и представит результат в виде вывода по тексту о наличии когнитивных искажений. Кроме того, XX в. дал нам существенные предпосылки к развитию психологии – труды Зигмунды Фрейда оказали большое влияние как на психоанализ, так и на людей, а после Второй мировой войны возникла проблема психологии посттравматического стрессового расстройства (ПТСР). В работе [1] показан тренд увеличения степени выраженности известных в когнитивно-поведенческой терапии когнитивных искажений в доступных публично текстах. В работах [2] и [3] показана выраженная связь между проявлением как эмоциональной составляющей, так и когнитивных искажений с движениями курсов криптовалют, что отражает степень воздействия манипулятивных воздействий на социум. Все это указывает на необходимость изучения психологии и степени влияния текстовых манипуляций на общество в целом, а также поиска методов противодействия этим манипуляциям. Сейчас в связи с доступностью большого количества информации психологические исследования стали открыты широкой общественности, благодаря чему люди начали обращать внимание на свое психологическое здоровье и, как следствие, чаще обращаться за помощью к специалистам. Таким образом, нагрузка на врачей возросла, а принципы диагностики остались прежними. Мы предлагаем алгоритм, который за счет анализа текста на наличие в них когнитивных искажений мог бы помочь в анализе публикаций в интернете и средах электронной коммуникации (мессенджерах), выявляя соответствующие манипулятивные воздействия, содержащиеся в этих текстах. Указанный алгоритм мог бы применяться для мониторинга общественного здоровья на основе публичных коммуникаций, как показано в работе [1], а также для персональной индикации потенциальных манипуляций в групповых коммуникациях (мессенджерах). Также в этой ситуации есть способ помочь

психотерапевту, создав ассистента, который, разговаривая с пациентом, проводил бы анализ когнитивных искажений в его речи и на основе полученных данных помогал бы специалисту ставить верный диагноз и отслеживать эффективность терапии, как предложено в работе [4]. В нашей работе мы представляем указанный алгоритм и проводим поиск и анализ оптимальных параметров для повышения его эффективности.

Методы и алгоритмы: Чтобы решить указанную задачу, мы создали набор словарей на английском языке, которые содержат в себе ключевые слова и фразы, характерные для каждого из выбранных нами когнитивных искажений на основе работы [1]. Далее описанный в работе [2] алгоритм был использован для анализа текста на предмет выраженности в нем соответствующих искажений. Он разбивает входные предложения на n-граммы и с учетом приоритетности, определяемой как порядок n-граммы N, проверяет их наличие в соответствующих словарях. Учет приоритетности представляет собой правило, по которому n-граммы меньшего размера, отвечающие одному когнитивному искажению и входящие в n-грамму большего размера, отвечающую другому когнитивному искажению, теряют свой вес, и выражение однозначно характеризуется когнитивным искажением большей n-граммы. После этого анализа алгоритм показывает коэффициент выраженности каждого искажения в тексте от 0 до 1. Затем мы делаем бинарную классификацию с порогами, установленными с шагом 0.1 и сравниваем полученные результаты при помощи F метрики с переменным параметром β с найденными по нашей тематике наборами данных.

Результаты и выводы: Мы выяснили, что лучше и универсальнее всего себя показывает среднее по всем искажениям с порогом 0.1 и F метрикой с параметром 0.1, что обусловлено наличием в датасетах большего числа текстов, содержащих искажения, чем нейтральных выражений. До порога 0.5 для F метрики с параметрами 0.1–0.5 также хорошо себя показывают искажения “negative statements”, “positive statements”, “rude statements” и “should statements”, что обусловлено неравномерным распределением выраженности когнитивных искажений в исследуемых наборах данных. С результатами можно ознакомиться по ссылке: https://github.com/AnnaArinicheva/CognitveProject.git

Diagnosis of cognitive distortions in public, group, and personal text communications

Arinicheva A.[1]*, Kolonin A.[1, 2]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Aigents, Novosibirsk, Russia*

* *a.arinicheva@g.nsu.ru*

Key words: cognitive distortions; search for optimal parameters; data analysis

Motivation and Aim: The enhancement of the average person's quality of life, the unstable global situation, and the broad availability of internet resources have primarily led to a sharp increase in the volume of information that people consume daily. This includes both verified facts and false claims, among which are manipulative texts aimed at deliberately distorting the world view and the mental state of their consumers. The virtual space exacerbates the difficulty of distinguishing authentic information from

falsification due to the absence of the ability to observe facial expressions, gestures, and intonation, which are characteristic of personal communication. In such situations, an assistant who can analyze the content of an article and present the results as an analysis of the text for cognitive distortions can help those faced with dubious news. Moreover, the 20th century provided significant groundwork for the development of psychology – the works of Sigmund Freud had a major impact on both psychoanalysis and people, and the issue of post-traumatic stress disorder (PTSD) emerged after World War II. Study [1] showed a trend of increasing expression of cognitive distortions known in cognitive-behavioral therapy in publicly available texts. Studies [2] and [3] demonstrated a pronounced correlation between the manifestation of both emotional components and cognitive distortions with the movements of cryptocurrency rates, reflecting the degree of manipulative impacts on society. All this indicates the necessity to study psychology and the degree of influence of textual manipulations on society as a whole, as well as searching for methods to counter these manipulations. Currently, with the availability of a large amount of information, psychological research has become accessible to the general public, as a result, people have started paying attention to their psychological health and, consequently, more often seek help from specialists. Thus, the burden on doctors has increased, while the principles of diagnosis have remained the same. We propose an algorithm that, through the analysis of text for cognitive distortions, could help in analyzing publications on the Internet and in electronic communication environments (messengers), identifying corresponding manipulative impacts contained in these texts. The specified algorithm could be used both for monitoring public health based on public communications, as shown in study [1], and for personal indication of potential manipulations in group communications (in messengers). Also, in this situation, there is a way to help the psychotherapist by creating an assistant who, by conversing with the patient, would analyze cognitive distortions in their speech and based on this data help the specialist make accurate diagnoses and monitor the effectiveness of therapy, as proposed in study [4]. In this work, we present the mentioned algorithm and conduct a search and analysis of optimal parameters to enhance its effectiveness.

Methods and Algorithms: Initially, we created a set of dictionaries in English that contain key words and phrases characteristic of each of the cognitive distortions we selected based on study [1]. Next, the algorithm described in study [2] was used to analyze the text for the degree of expression of the corresponding distortions. It breaks down input sentences into n-grams and, considering the priority determined as the order of n-gram N, checks their presence in the corresponding dictionaries. Priority consideration is a rule whereby smaller-sized n-grams corresponding to one cognitive distortion and included in a larger n-gram corresponding to another cognitive distortion lose their weight, and the expression is unequivocally characterized by the cognitive distortion of the larger n-gram. After this analysis, the algorithm displays the degree of expression of each distortion in the text from 0 to 1. We then perform binary classification with thresholds set at increments of 0.1, and compare the results using the F metric with a variable parameter β against datasets found in our theme.

Results and Conclusion: We have discovered that the average across all distortions with a threshold of 0.1 and an F metric with a parameter of 0.1 shows the best and most universal performance, which is due to the presence in the datasets of a larger number of texts containing distortions than neutral expressions. Up to the threshold of 0.5 for F metrics with parameters from 0.1 to 0.5, distortions such as “negative statements”,

“positive statements”, “rude statements”, and “should statements” also perform well, which is due to the uneven distribution of the expression of cognitive distortions in the studied datasets. You can view the results at the following link: https://github.com/AnnaArinicheva/CognitveProject.git

Список литературы/References

1. Bollen J., Thij M., Breithaupt F., Barron A., Rutter L., Lorenzo-Luaces L., Scheffer M. Historical language records reveal a surge of cognitive distortions in recent decades. *PNAS.* 2021;118(30):e2102061118. doi 10.1073/pnas.2102061118
2. Raheman A., Kolonin A., Fridkins I., Ansari I., Vishwas M. Social Media Sentiment Analysis for Cryptocurrency Market Prediction. *arXiv*. 2022. doi 10.48550/arXiv.2204.10185
3. Kolonin A., Raheman A., Vishwas M., Ansari I., Pinzon J., Ho A. Causal Analysis of Generic Time Series Data Applied for Market Prediction. *arXiv*. 2022. doi 10.48550/arXiv.2204.12928
4. Kolonin A., Kurpatov A., Molchanov A., Averyanov. Cognitive Architecture for Decision-Making Based on Brain Principles Programming. *Procedia Comput Sci.* 2022;213:180-189. doi 10.1016/j.procs.2022.11.054

doi 10.18699/bgrs2024-10-02

Выделение и анализ активных ансамблей нервных клеток, ассоциированных с паттернами поведения мыши

Варехина А.В.[1]*, Иванченко М.В.[1], Сотсков В.П.[2, 3], Кривоносов М.И.[1], Анохин К.В.[4]

[1] *Институт информационных технологий, математики и механики, Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия*

[2] *Центр междисциплинарных исследований в области биологии, Колледж де Франс, Париж, Франция*

[3] *Институт биологии, Высшая нормальная школа, Париж, Франция*

[4] *Институт перспективных исследований мозга, Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

* *varekhina@unn.ru*

Ключевые слова: кальциевый имиджинг; активность нервных клеток; анализ данных

Мотивация и цель: Кальциевый имиджинг – метод, позволяющий регистрировать кальциевую динамику в нервных клетках не только в культурах, но и непосредственно в мозге объекта исследования. Полученные с помощью этого метода результаты могут быть использованы как в изучении влияния определенных физиологических состояний на работу нервных клеток, так и при анализе механизмов взаимодействий отдельных клеточных ансамблей между собой [1]. В этом исследовании мы изучали взаимосвязь между поведением животного и сетевой кальциевой активностью нервных клеток.

Методы и алгоритмы: Нервные импульсы регистрировались с помощью установленного на голове минископа NVista HD. Мышь с минископом помещали в кольцевой трек, предварительно очищенный от посторонних запахов. Вдоль взаимно перпендикулярных диаметров дорожки расположены четыре метки, позволяющие животному делать какие-либо выводы о своем текущем положении. В каждый момент времени в ходе эксперимента на видео фиксировались координаты положения мыши [2, 3]. Мы предлагаем алгоритм, позволяющий выделять графы взаимодействия нервных клеток и анализировать их кальциевую динамику в течение времени. Сначала была проведена разметка поведенческих видео и выделены следующие типы деятельности мыши: сидит на месте, бежит, умывается, смотрит наружу или внутрь кольцевого трека, стойка наружу или внутрь кольцевого трека. Для моментов времени, когда нельзя было однозначно сказать, что происходит на видео, был введен класс «не определено». Далее были построены графы взаимодействия клеток в течение времени с парами активных клеток. По построенному графу и парам взаимодействий клеток за все время видео были получены группы из трех последовательно загорающихся клеток – тройки активных нейронов. В рассмотрении участвовали только те тройки клеток, внутри которых были только две связи между клетками и строгое отсутствие третьей (рис. 1), что позволяет установить предполагаемое распространение сигнала. Для того чтобы выделить взаимосвязь между определенным паттерном поведения и активностью клеток в тройках, был применен критерий хи-квадрат с поправкой на множественные сравнения на множество троек. Для демонстрации значимости

связи между комбинациями из трех клеток и поведением мыши была исследована работа алгоритма на графах, построенных на рандомизированной последовательности кальциевых событий [4].

Результаты: Таким образом, были выделены статистически значимые тройки клеток, ассоциированные с паттернами поведения, и восстановлены воспроизводимые временные соотношения в их активации (рис. 2). В графе с 562 вершинами в течение 15-минутного видео было выделено 1.6 млн троек для оригинального и 2.3 млн троек для рандомизированного графов. Такое большое число найденных троек связано с низким числом их повторений за время эксперимента. В связи с этим было предложено отобрать только тройки, наблюдающиеся не менее 5 раз. Такой выбор порогового значения обусловлен тем, что в рандомизированном графе отсутствуют статистически значимые тройки, ассоциированные с поведением мыши и наблюдавшиеся более 4 раз. При этом в реконструированном графе на основе кальциевой активности нейронов было обнаружено 2064 тройки, наблюдавшихся более 4 раз (рис. 2), а статистически значимых из них – 433.

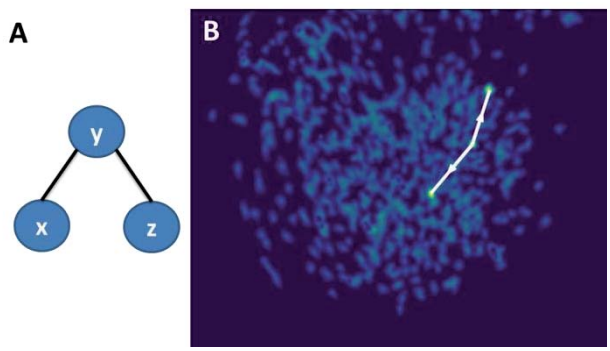


Рис. 1. A – конфигурация искомых троек взаимодействующих клеток. B – общий внешний вид тройки взаимодействующих клеток. Стрелками указана последовательность активации клеток. Желто-зеленые участки – клетки в пределах тройки, синие участки – клетки, не участвующие в рассмотрении

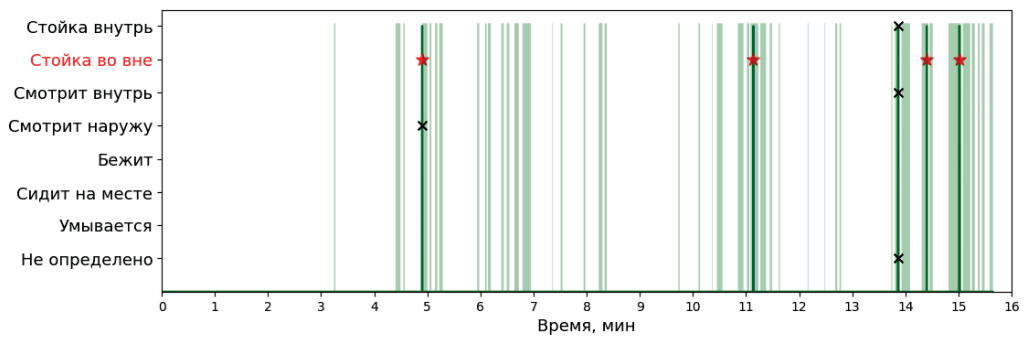


Рис. 2. Время горения тройки взаимодействующих клеток. Темно-зеленым отмечены периоды активности тройки клеток, светло-зеленым – периоды «лучшего» класса, во время которого дольше всех были активны клетки в пределах тройки. Название этого класса выделено красным цветом. Красными звездочками отмечены моменты активности тройки в пределах лучшего класса, зелеными крестами – моменты активности тройки в пределах других классов

Выводы: Исследование показало, что существуют статистически значимые ансамбли клеток, активные только во время определенной деятельности, причем их состав и частота активности отличались от таковых в рандомизированном графе. Это говорит о том, что есть закономерность между передачами сигнала в клетках. Полученные результаты можно использовать в предсказании паттернов поведения по кальциевой динамике нервных клеток и разработке неинвазивных интерфейсов мозг-компьютер.
Финансирование: Исследование выполнено при финансовой поддержке «Центра фотоники», финансируемого Министерством науки и высшего образования Российской Федерации (контракт № 075-15-2022-293).

Identification and analysis of active neural cell ensembles associated with mice behavioral patterns

Varekhina A.V.[1]*, Ivanchenko M.V.[1], Sotskov V.P.[2, 3], Krivonosov M.I.[1], Anokhin K.V.[4]

[1] *Institute of Information, Technology, Mathematics and Mechanics, Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia*
[2] *Center of Interdisciplinary Research in Biology, Collège de France, Paris, France*
[3] *Institute of Biology, École Normale Supérieure, Paris, France*
[4] *Institute for Advanced Brain Studies, Lomonosov Moscow State University, Moscow, Russia*
* *varekhina@unn.ru*

Key words: calcium imaging; neural activity; data analysis

Motivation and Aim: Calcium imaging is a method that allows recording calcium dynamics in neural cells not only in cultures, but also directly in the brain of the object of study. The results obtained using this method can be used both in the study of the influence of certain physiological states on the work of neural cells, and in the analysis of the mechanisms of interactions of individual cell ensembles with each other [1]. In this study, we examined the relationship between animal behavior and the network calcium activity of nerve cells.
Methods and Algorithms: Neural impulses were recorded using an NVista HD head-mounted miniscope. The mouse with the miniscope was placed in a ring track, previously cleaned of foreign odors. Four marks are placed along the mutually perpendicular diameters of the track, allowing the animal to draw conclusions about its current position. At each time point during the experiment, the coordinates of the mouse position were recorded in the video [2, 3]. We propose an algorithm that allows us to isolate graphs of the interaction of neural cells and analyze their calcium dynamics over time. First, behavioral videos were labeled and the following types of mouse activity were distinguished: sitting, running, grooming, looking outside or inside the ring track, standing on the outside or inside edge of the ring track. For times when it was impossible to say certainly what was going on in the video, the label “undefined” was introduced. Next, graphs of the interaction of cells over time with pairs of active cells were constructed. According to the constructed graph and pairs of cell interactions, groups of sequentially igniting cells (triples of active neurons) were obtained. Only those triples of cells were considered, within which there were only two connections between cells and

a strict absence of a third (Fig. 1), which allows to establish the supposed signal propagation. In order to isolate the relationship between a certain pattern of behavior and the activity of cells in triples, the chi-square test was applied, adjusted for multiple comparisons for multiple triples. To demonstrate the significance of the relationship between combinations of triples cells and mouse behavior, the work of the algorithm on graphs based on a randomized sequence of calcium events was studied [4].

Results: Thus, statistically significant triples of cells associated with behavior patterns were identified, and reproducible temporal relations in their activation were restored (Fig. 2). In a graph with 562 vertices during a 15-minute video, 1.6 million triples were identified for the original and 2.3 million triples for the randomized graphs. Such a large amount of found triples is associated with a low number of their repetitions during the experiment. In this regard, it was suggested to select only triples observed at least 5 times. This choice of threshold value is due to the fact that there are no statistically significant triples associated with mouse behavior and observed more than 4 times in the randomized graph. At the same time, 2064 triples were found in the reconstructed graph based on the calcium activity of neurons, which were observed more than 4 times, and 433 of them were statistically significant.

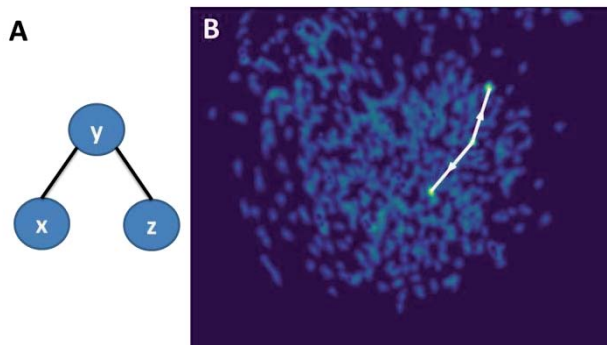


Fig. 1. A – Configuration of the desired triples of interacting cells. **B –** General appearance of the triple of interacting cells. The arrows indicate the cell activation sequence. Yellow-green areas signify cells within the three, while blue areas are cells not taken into consideration

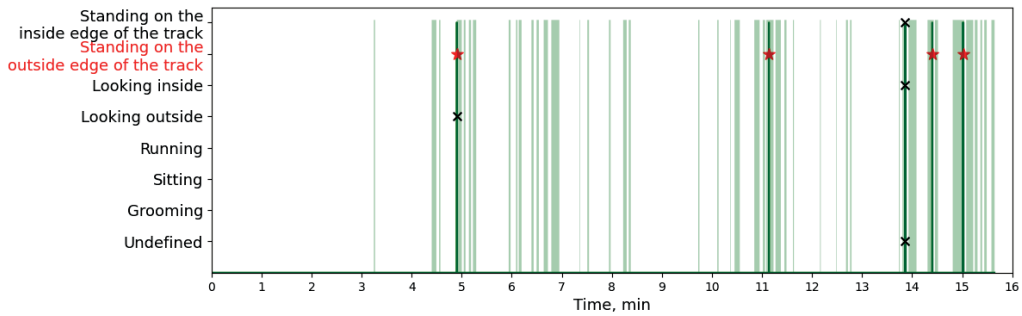


Fig. 2. Activity time of the interacting cells triples. Dark green marked the periods of activity of the three cells, light green marked periods of the "best" label, during which the cells within the three were active for the longest time. The name of this label is highlighted in red. Red asterisks mark the moments of activity of the three within the best label, green crosses mark the moments of activity of the three within other labels

Conclusion: The study showed that there are statistically significant ensembles of cells that are active only during a certain activity, and their composition and frequency of activity differed from those in the randomized graph. This suggests that there is a pattern between signal transmissions in cells. The results obtained can be used in the prediction of behavior patterns by the calcium dynamics of nerve cells and the development of non-invasive brain-computer interfaces.

Funding: The study is supported by the "Center of Photonics" funded by the Ministry of Science and Higher Education of the Russian Federation (contract No. 075-15-2022-293).

Список литературы/References

1. Robbins M., Christensen C.N., Kaminski C.F., Zlatic M. Calcium imaging analysis – how far have we come? *F1000Res.* 2021;10:258
2. Sotskov V.P., Plusnin V.V., Pospelov N.A., Anokhin K.V. The Rapid Formation of CA1 Hippocampal Cognitive Map in Mice Exploring a Novel Environment. In: Advances in Cognitive Research, Artificial Intelligence and Neuroinformatics: Proceedings of the 9th International Conference on Cognitive Sciences, October 10-16, 2020, Moscow, Russia. Springer International Publishing, 2021;452-457
3. Sotskov V.P., Pospelov N.A., Plusnin V.V., Anokhin K.V. Calcium imaging reveals fast tuning dynamics of hippocampal place cells and CA1 population activity during free exploration task in mice. *Int J Mol Sci.* 2022;23(2):638
4. Hanhijärvi S., Garriga G.C., Puolamäki K. Randomization techniques for graphs. In: Proceedings of the 2009 SIAM international conference on data mining. Society for Industrial and Applied Mathematics, 2009;780-791

doi 10.18699/bgrs2024-10-04

Влияние изменения экспрессии гена *limk1* в дофаминергических, серотонинергических и fruitless нейронах на обучение и забывание

Заломаева Е.С.[1, 2]*, Егозова Е.С.[1], Медведева А.В.[2], Журавлёв А.В.[2], Никитина Е.А.[1, 2]

[1] *Российский государственный педагогический университет им. А.И. Герцена, Санкт-Петербург, Россия*

[2] *Институт физиологии им. И.П. Павлова РАН, Санкт-Петербург, Россия*

* *Zalomaeva.E@yandex.ru*

Ключевые слова: *limk1*; *Drosophila melanogaster*; память; обучение; забывание

Мотивация и цель: В настоящее время одной из актуальных проблем современной нейробиологии является понимание этиологии и прогрессирования различных нейродегенеративных заболеваний. Это группа в основном медленно прогрессирующих заболеваний нервной системы, характеризующихся стремительной гибелью нейронов. Среди наиболее распространенных нейродегенеративных заболеваний выделяют болезнь Альцгеймера, Паркинсона, Хантингтона и др. Данные заболевания могут быть результатом сложного взаимодействия неблагоприятных внешних факторов, а также индивидуальных особенностей генома, предрасполагающих к развитию болезни. На сегодняшний день не существует лекарств и методов лечения, которые бы смогли остановить развитие патологии. Даже поддержание пациента на поздних стадиях заболевания в ясном состоянии сознания представляется практически невозможным. Поэтому современная наука стремится к решению вопроса о причинах возникновения и прогрессирования данных заболеваний, чтобы в перспективе излечивать их еще на ранних стадиях или предупреждать их возникновение. Для лучшего понимания патогенеза нейродегенеративных заболеваний необходимо изучить основу механизмов и процессов памяти, таких как обучение и забывание [1]. Но поскольку сохранение памяти является результатом обоих процессов – обучения и забывания, возникает вопрос: как понять, что в действительности является причиной «когнитивной патологии» – дефект обучения и консолидации или же дефектность активного врожденного механизма забывания. Активное забывание регулируется сигнальным каскадом ремоделирования актина, ключевой фермент которого LIMK1 [2]. Изменения экспрессии гена *limk1* могут приводить к нейрокогнитивным патологиям. Цель исследования – изучение роли гена *limk1* в процессах обучения и забывания у *Drosophila melanogaster*.

Методы и алгоритмы: Исследование проводили на самцах линий дрозофилы с нейроспецифическим подавлением и активацией гена *limk1* в нейронах fruitless, а также в дофаминергических и серотонинергических нейронах. В качестве контроля использовали линию без изменения экспрессии гена *limk1*. Для изменения экспрессии применяли систему скрещивания Gal4-UAS. Для оценки способности к обучению и формированию памяти использовали методику условно-рефлекторного подавления ухаживания. Пятисуточного девственного

самца помещали в специальную камеру вместе с оплодотворенной самкой линии дикого типа *CantonS* на 30 минут. Самцов тестировали через разные интервалы времени после тренировки. В качестве контроля использовали самцов, не имеющих опыта полового поведения. За поведением обученных и наивных самцов наблюдали в течение 300 секунд, фиксируя в специальной программе время начала и окончания отдельных элементов [3]. Элементы могли быть связаны с ухаживанием (ориентация, вибрация крыла, постукивание, лизание, попытка копуляции) или не связаны с ним (активность, прининг, покой). Для оценки эффективности обучения вычисляли индекс обучения (ИО) по следующей формуле:

$$ИО = \left[\frac{ИУн - ИУт}{ИУн}\right] \times 100\% = \left(1 - \frac{ИУт}{ИУн}\right) \times 100\%,$$

где ИО – индекс обучения; ИУн – средний индекс ухаживания для самцов, не имеющих опыта полового поведения; ИУт – средний индекс ухаживания для обученных самцов.

Для оценки активности процессов забывания провели анализ скорости снижения ИО на коротких временных интервалах (0, 15, 30, 60 минут после тренировки) и спустя 24 часа, когда ИО может оставаться высоким при нарушении активного забывания. В каждой группе тестировали по 20 пар мух. Для статистической обработки данных использовали двусторонний тест рандомизации ($p \leq 0.05$).

Результаты: При анализе ИО линий с изменением экспрессии гена *limk1* в нейронах fruitless выявили, что мухи контрольной линии оказались способны сохранять память в течение 15 минут. Подавление *limk1* в нейронах fruitless восстанавливает память на интервале 30–60 минут, таким образом снижая процесс забывания. У линий с активацией *limk1* в нейронах fruitless в большинстве временных точек значения ИО были низкими, а спустя 60 минут после тренировки ИО достоверно отличался от линии с подавлением *limk1*. Спустя 24 часа после тренировки ИО всех трех линий имели схожие значения. У контрольной линии ИО спустя 24 часа был достоверно ниже, чем сразу после тренировки, а у линии с активацией *limk1* – достоверно ниже, чем спустя 15 минут после тренировки.

При анализе ИО линий с изменением экспрессии гена *limk1* в дофаминергических и серотонинергических нейронах выявили, что контрольная линия проявляет способность к обучению и сохранению памяти вплоть до 30 минут. При этом спустя 24 часа память у данной линии резко нарушена, и ИО принимает отрицательное значение. Подавление *limk1* в дофаминергических и серотонинергических нейронах снижает ИО, которые статистически не отличаются от нуля уже сразу после тренировки. Мухи линии с активацией *limk1* в дофаминергических и серотонинергических нейронах способны к обучению, однако спустя 24 часа после тренировки ИО также принял отрицательное значение.Кривые забывания у мух всех трех линий сходны: на протяжении 60 минут ИО сохраняется на относительно постоянном уровне, а через 24 часа после тренировки приобретает отрицательные значения.

Выводы: Таким образом, полученные данные говорят о вовлеченности гена *limk1* в процессы обучения и забывания. При этом характер влияния нейроспецифического изменения экспрессии гена *limk1* на динамику ИО определяется конкретным типом нейронов. Полученные результаты открывают перспективы дальнейших исследований с целью выявления путей целенаправленного

терапевтического воздействия на белки и гены, вовлеченные в развитие нейродегенерации.
Финансирование: Исследование поддержано средствами федерального бюджета в рамках государственного задания ФГБУН Институт физиологии им. И.П. Павлова РАН (№ 1021062411629-7-3.1.4).

The effect of changes *limk1* gene expression in dopaminergic, serotonergic and fruitless neurons on learning and forgetting

Zalomaeva E.S.[1, 2]*, Egozova E.S.[1], Medvedeva A.V.[2], Zhuravlev A.V.[2], Nikitina E.A.[1, 2]

[1] *Herzen State Pedagogical University of Russia, St. Petersburg, Russia*

[2] *Pavlov Institute of Physiology RAS, St. Petersburg, Russia*

* *Zalomaeva.E@yandex.ru*

Key words: *limk1*; *Drosophila melanogaster*; memory; learning; forgetting

Motivation and Aim: Nowadays one of the topical problems of neurobiology is the research of the etiology and progression of different neurodegenerative diseases. This is a group of diseases of the nervous system that are characterized by rapid death of neurons and slow progression. The most common neurodegenerative diseases are Alzheimer's disease, Parkinson's disease, Huntington's disease. The causes of these disorders may be a complex interaction of unfavorable external factors, as well as individual features of the genome predisposing to the development of the disease. Today there are no medications or treatments which can to stop the development of the pathology. Even maintaining a patient in the later stages of the disease in a clear state of consciousness seems almost impossible. Therefore modern science tries find the causes of the onset and progression these diseases. In perspective, it will help to cure these diseases at an early stage or to prevent their onset. To better understand the pathogenesis of neurodegenerative diseases, it is necessary to study the underlying mechanisms and processes of memory, such as learning and forgetting [1]. But since the preservation of memory is the result of both processes – learning and forgetting – there is a question: how to understand what is the cause of “cognitive pathology” – a defect in learning and consolidation or a defect in the active innate mechanism of forgetting. Active forgetting is regulated by the actin remodeling signaling cascade, the key enzyme is LIMK1 [2]. Changes in the expression of the *limk1* gene can lead to neurocognitive pathologies. The aim of the research was to study the role of the *limk1* gene in the processes of learning and forgetting in *Drosophila melanogaster*.
Methods and Algorithms: We conducted a study on males drosophila lines with neurospecific suppression and activation of the *limk1* gene in fruitless neurons, as well as in dopaminergic and serotonergic neurons. The control line was used without changing the expression of the *limk1* gene. The Gal4-UAS crossing system was used to change the expression. Conditioned courtship suppression paradigm was used to asses learning ability and memory formation. A 5-day-old virgin male was placed in a special box together with a fertilized wild-type female for 30 minutes. The males were tested at different time intervals after training. Males with no experience of sexual behavior were used as controls. The behavior of trained and naive males was observed for 300 seconds,

recording the start and end times of individual elements in a special program [3]. The elements could be related to courtship (orientation, wing vibration, tapping, licking, attempt at copulation), or unrelated to it (activity, pruning, rest). To assess the effectiveness of training the learning index is calculated. Formula of learning index:

$$LI = \left[\frac{CIn - CI_T}{CIn}\right] \times 100\% = \left(1 - \frac{CI_T}{CIn}\right) \times 100\%,$$

where LI – is the learning index; CIn – is the middle Courtship index for control males; CI_T – is the middle Courtship index for males after training.

To assess the activity of the forgetting processes, an analysis of the rate of decrease in LI was performed at short time intervals (0, 15, 30, 60 minutes after training) and 24 hours later, when LI may remain high with a violation of active forgetting. 20 pairs of flies were tested in each group. A two-way randomization test ($p \leq 0.05$) was used for statistical data processing.

Results: When analyzing LI stock with a change in the expression of the *limk1* gene in fruitless neurons, it was revealed that flies of the control stock were able to retain memory for 15 minutes. Suppression of *limk1* in fruitless neurons restores memory in the range of 30–60 minutes, thus reducing the process of forgetting. In the *limk1* activation stock in fruitless neurons, LI values were low at most points, and 60 minutes after training, LI significantly differed from the stock with *limk1* suppression. 24 hours after training, the LI of all three stocks have similar values. The control stock has significantly lower LI after 24 hours than immediately after training, and the stock with *limk1* activation has significantly lower LI than 15 minutes after training.

When analyzing LI stocks with changes in the expression of the *limk1* gene in dopaminergic and serotonergic neurons, it was revealed that the control stock showed the ability to learn and retain memory up to 30 minutes. At the same time, 24 hours later, the memory of this stockis sharply disrupted, and LI has taken a negative value. Suppression of *limk1* in dopaminergic and serotonergic neurons reduces LI, which statistically do not differ from zero immediately after exercise. A stock with *limk1* activation in dopaminergic and serotonergic neurons showed the ability to learn, but 24 hours after training, LI also took a negative value. The forgetting curves in flies of all three stocks show a similar tendency. For 60 minutes, the LI remains at a relatively constant level, and 24 hours after training it acquires negative values.

Conclusion: Thus, the data obtained indicate the unconditional involvement of the *limk1* gene in the processes of learning and forgetting. At the same time, the nature of the effect of a neurospecific change in the *limk1* gene on the dynamics of LI is determined by a specific type of neurons. The results obtained require further research in order to identify ways of targeted therapeutic effects on proteins and genes involved in the development of neurodegeneration.

Funding: The study is supported by the State funding allocated to the Pavlov Institute of Physiology, Russian Academy of Sciences (No. 1021062411629-7-3.1.4).

Список литературы/References

1. Medina C., de la Fuente V., Dieck T.S. et al. LIMK, Cofilin 1 and actin dynamics involvement in fear memory processing. *Neurobiol Learn Mem*. 2020;173:107275. doi 10.1016/j.nlm.2020.107275
2. Davis R.L., Zhong Y. The biology of forgetting – a perspective. *Neuron*. 2017;95(3):490-503. doi 10.1016/j.neuron.2017.05.039
3. Kamyshev N.G., Iliadi K.G., Bragina J.V. *Drosophila* conditioned courtship: two ways of testing memory. *Learn Mem*. 1999;6(1):1-20

doi 10.18699/bgrs2024-10-05

ЭЭГ реакции при распознавании оценочных предложений про себя и других у подростков в сравнении со взрослыми

Злая С.[1]*, Сапрыгин А.[2, 3], Лебедкин Д.[1, 3], Степанова В.[4], Савостьянов А.[1, 2, 3]

[1] *Новосибирский государственный университет, Гуманитарный институт, Новосибирск, Россия*

[2] *Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия*

[3] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[4] *Областной центр информационных технологий, Новосибирск, Россия*

* *s.zlaya@g.nsu.ru*

Ключевые слова: самореференция; ЭЭГ; вызванные потенциалы головного мозга; семантическое и грамматическое распознавание речи

Мотивация и цель: Под самореференцией понимается отнесение человеком какой-либо информации к самому себе [1]. Самоотнесенные высказывания человек воспринимает как сообщения о нем самом, в противоположность к высказываниям о других людях или вещах. Исследование процессов, связанных с распознаванием информации о себе и других людях, вызывает большой интерес у широкого круга специалистов, включая психологов, нейрофизиологов и лингвистов. Актуальность данной темы связана с решением фундаментальных вопросов о нейрофизиологических механизмах формирования человеческой личности. Кроме того, самореферентные процессы могут быть нарушены при разных психических заболеваниях [2]. Ранее было показано, что анализ ЭЭГ реакций, регистрируемых в условиях распознавания письменных предложений про себя и других людей, позволяет исследовать нейрофизиологические механизмы, лежащие в основе самореференции [3]. В нашем новом исследовании мы проверяем гипотезу, что особенности распознавания самоотнесенной информации могут в существенной степени зависеть от возраста обследуемых людей. В частности, можно предположить, что у подростков при распознавании самоотнесенных письменных предложений могут быть задействованы другие, по сравнению со взрослыми, области коры головного мозга, и, соответственно, нейрофизиологические процессы, лежащие в основе самореференции, отличаются между подростками и взрослыми.

Цель: Изучить и сравнить реакции здоровых людей разного возраста на самоотнесенные и не самоотнесенные эмоционально окрашенные и не окрашенные предложения. На основании полученных данных сформировать вывод об участии различных мозговых структур в обработке самоотнесенной информации у взрослых и подростков.

Методы и алгоритмы: *1. Участники*. В эксперименте было обследовано 20 школьников в возрасте от 12 до 14 лет и 20 студентов в возрасте от 18 до 24 лет. Все студенты подписывали согласие на участие в обследовании лично, за подростков согласие давали их родители. *2. Экспериментальное задание.* Каждому участнику в случайном порядке предъявлялось 144 письменных предложения на русском языке, половина из которых содержала грамматическую ошибку. Задачей участника было определить, является ли предложение грамматически правильным или нет. Кроме того, все предложения относились к одной из шести семантических

групп, о чем участнику обследования не сообщалось. Предложения содержали: 1) нейтральную оценку самого участника, 2) негативную оценку самого участника, 3) положительную оценку самого участника, 4) нейтральную оценку других людей, 5) негативную оценку других людей, 6) положительную оценку других людей. *3. Регистрация ЭЭГ*. ЭЭГ с расстановкой меток событий записывалась по 128 каналам, расположенным по международной схеме 5–10 %, при помощи усилителя NVX-136, Россия. Артефакты удалялись при помощи анализа независимых компонент (ICA). В качестве показателей мозговой активности, ассоциированных с распознаванием предложений, были выбраны амплитуды пиков вызванных потенциалов P300 и N400.

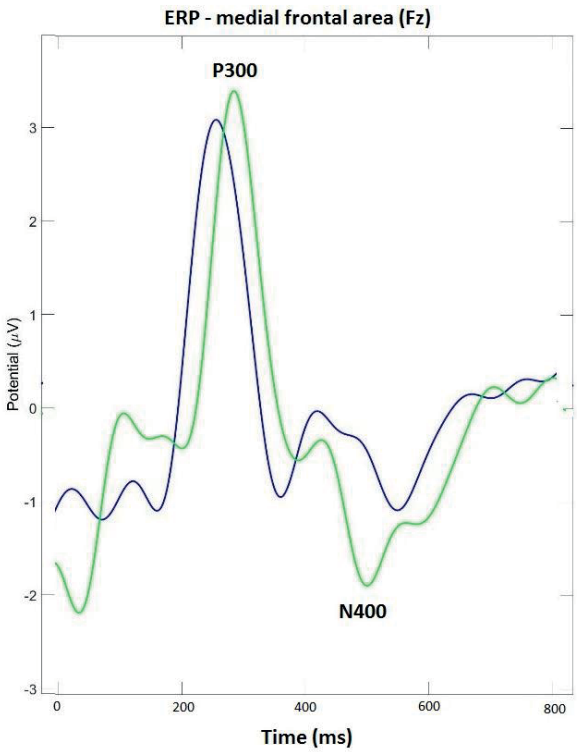


Рис. 1. График амплитудно-временной зависимости событийных потенциалов (пики P300 и N400) у взрослых (синяя линия) и подростков (зеленая линия)

Результаты: На амплитудно-временном графике вызванных потенциалов были идентифицированы пики P300 и N400 (рис. 1). Пик P300 отражает уровень субъективного внимания к стимулу, а пик N400 отражает процессы семантической обработки информации. Были выявлены достоверные различия в амплитудах ERP пиков на предложения, относящиеся к разным семантическим категориям. У всех испытуемых, независимо от пола и возраста, большая амплитуда пика P300 наблюдалась на предложения про себя, чем на предложения про других людей, вне зависимости от их эмоциональной модальности. Различия в мозговых ответах между подростками и взрослыми касались в первую очередь корковой топографии ответов на самоотнесенные и несамоотнесенные предложения. У взрослых различия в ответах на разные категории предложения наблюдались в зоне Брока (левая лобно-височная область коры), а у подростков – в зоне Вернике (левая

теменно-височная область коры) (рис. 2 и 3). Кроме того, у подростков наибольшая амплитуда для всех рассмотренных пиков наблюдалась на эмоционально положительные предложения про самого себя, а у взрослых – на отрицательно окрашенные предложения о других людях.

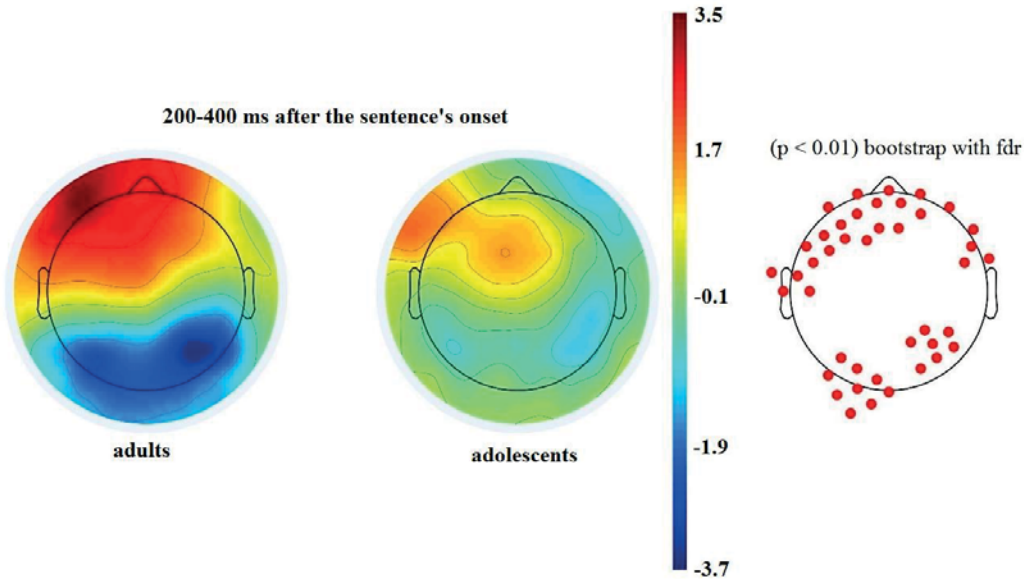


Рис. 2. Корковое распределение пиковой амплитуды P300 в течение 200–300 мс по поверхности коры у взрослых (слева) и подростков (справа)

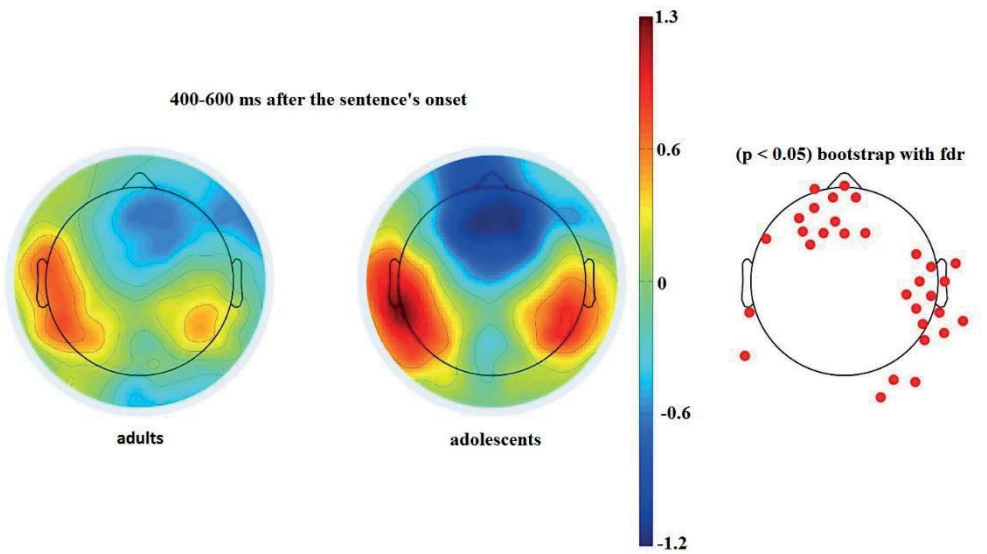


Рис. 3. Корковое распределение пиковой амплитуды N400 в интервале 400–600 мс по поверхности коры у взрослых (слева) и подростков (справа)

Выводы: Вызванные потенциалы головного мозга в условиях поиска грамматической ошибки в письменных предложениях про самого себя и других людей отражают возрастные особенности распознавания личностно окрашенной эмоциональной информации. Можно предположить, что различия между подростками и взрослыми касаются разной пропорциональной отнесенности между распознаванием синтаксических и семантических характеристик речи, что на ERP отражено в сдвиге максимальной амплитуды пиков от зоны Вернике у подростков, к зоне Брока – у взрослых.
Финансирование: Исследование поддержано за счет средств гранта РНФ № 22-15-00142 «фМРТ и ЭЭГ корреляты фокуса внимания на собственной персоне как фактора предрасположенности к аффективным расстройствам». Работа А.Е. Сапрыгина и А.Н. Савостьянова по разработке методик анализа нейрофизиологических данных финансировалась из средств бюджетного проекта ИЦиГ СО РАН № FWNR-2022-0020.

EEG reactions in recognizing evaluative sentences about oneself and others in adolescents compared to adults

Zlaya S.[1]*, Saprigyn A.[2, 3], Lebedkin D.[1, 3], Stepanova V.[4], Savostyanov A.[1, 2, 3]

[1] *Novosibirsk State University, Humanities Institute, Novosibirsk, Russia*

[2] *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[4] *Regional Information Technology Cenetr, Novosibirsk, Russia*

** s.zlaya@g.nsu.ru*

Key words: self-reference; EEG; evoked brain potentials; semantic and grammatical speech recognition; self- and non-self-relational sentences

Motivation and Aim: Self-reference refers to a person's attribution of any information to himself/herself [1]. A person perceives self-relational statements as messages about himself, as opposed to statements about other people or things. The study of processes related to the recognition of information about oneself and other people is of great interest to a wide range of specialists, including psychologists, neurophysiologists and linguists. The relevance of this topic is related to solving fundamental questions about the neurophysiological mechanisms of human personality formation. In addition, self-relational processes can be disrupted in various mental illnesses [2]. Previously, it was shown that the analysis of EEG reactions recorded in the conditions of recognition of written sentences about oneself and other people makes it possible to investigate the neurophysiological mechanisms underlying self-reference [3]. In our new investigation, we test the hypothesis that the features of recognizing self-relational information may significantly depend on the age of the people being investigated. Particularly, it can be assumed that in adolescents, when recognizing self-relational written sentences, other cortical zones may be involved, compared with adults, and accordingly, the neurophysiological processes underlying self-reference differ between adolescents and adults.
Purpose: To study and compare the reactions of healthy people of different ages to self-relational and non-self-relational emotionally colored and non-colored sentences. Based

on the data obtained, to form a conclusion about the participation of various brain structures in the processing of self-referential information in adults and adolescents.
Methods and Algorithms: *1. Participants*. In the experiment, 20 schoolchildren aged 12 to 14 years and 20 students aged 18 to 24 years were investigated. All students signed consent to participate in the survey personally, and their parents gave consent for the adolescents. *2. An experimental task*. Each participant was randomly presented with 144 written sentences in Russian, half of which contained a grammatical error. The participant's task was to determine whether the sentence was grammatically correct or not. In addition, all the sentences belonged to one of the six semantic groups, which was not reported to the survey participant. The proposals contained (1) a neutral assessment of the participant himself, (2) a negative assessment of the participant himself, (3) a positive assessment of the participant himself, (4) a neutral assessment of other people, (5) a negative assessment of other people, (6) a positive assessment of other people. *3. EEG registration*. The event-related EEG was recorded over 128 channels arranged according to the international 5–10 % scheme using an NVX-136 amplifier, Russia. Artifacts were removed using Independent Component Analysis (ICA). The peak amplitudes of the event-related potentials P300 and N400 were selected as indicators of brain activity associated with sentence recognition.

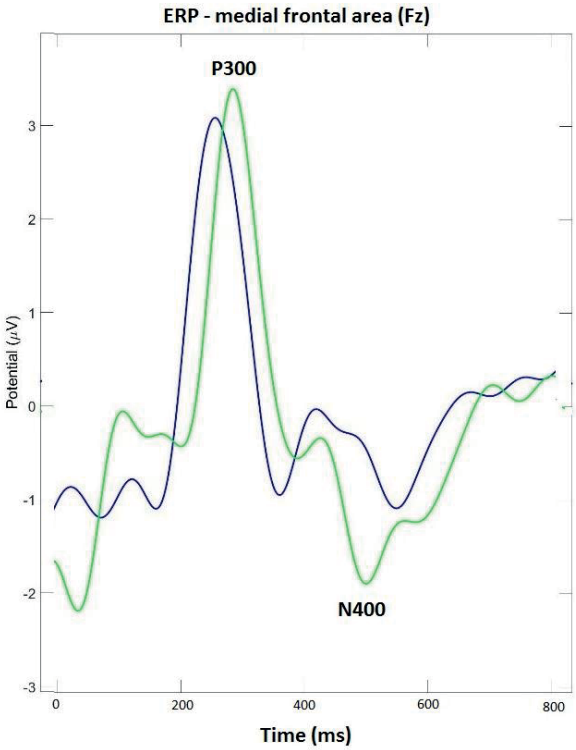


Fig. 1. The amplitude-time plot of event-related potentials (peaks P300 and N400) in adults (blue line) and adolescents (green line)

Results: Peaks P300 and N400 were identified on the amplitude-time plot of ERPs (Fig. 1). The peak of P300 reflects the level of subjective attention to the stimulus, the peak of N400 reflects the processes of semantic information processing. Significant

differences in the amplitudes of ERP peaks for sentences belonging to different semantic categories were revealed. In all subjects, regardless of gender and age, a greater amplitude of the P300 peak was observed for sentences about oneself than for sentences about other people, regardless of their emotional modality. The differences in brain responses between adolescents and adults concerned primarily the cortical topography of responses to self-related and non-self-related sentences. In adults, differences in responses to different categories of sentences were observed in the Broca's area (left frontal-temporal cortex), and in the adolescents – in the Wernicke's area (left parietal-temporal cortex) (Fig. 2 and 3). In addition, adolescents had the highest ERP's amplitude to emotionally positive sentences about themselves, and adults had the highest amplitude of reactions to negatively colored sentences about other people.

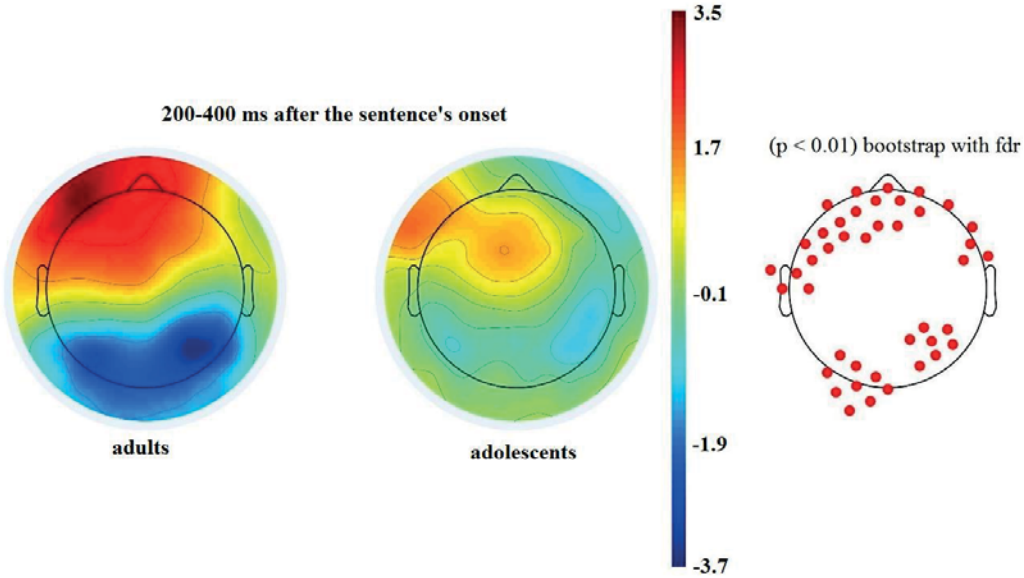


Fig. 2. The cortical distribution of peak P300 amplitude in 200–300 ms across the cortical surface for adults (left) and adolescents (right)

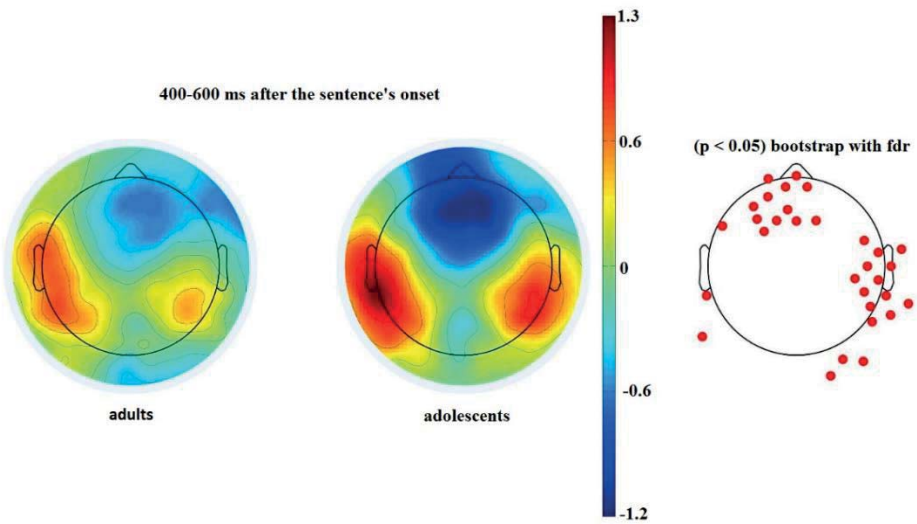


Fig. 3. The cortical distribution of peak N400 amplitude in 400-600 ms across the cortical surface for adults (left) and adolescents (right)

Conclusion: The brain's event-related potentials in the search for a grammatical error in written sentences about oneself and other people reflect the age-related features of recognizing personally colored emotional information. It can be assumed that the differences between adolescents and adults relate to the different proportional relationship between the recognition of syntactic and semantic characteristics of speech, which is reflected in ERP in the shift of the maximum amplitude of peaks from the Wernicke's area in adolescents to the Broca's area in adults.

Funding: The study was supported by the Russian Science Foundation Grant No. 22-15-00142 "fMRI and EEG correlates of self-focus as a factor of predisposition to affective disorders". The work of Saprigyn A.E. and Savostyanov A.N. on the development of methods for analyzing neurophysiological data was supported by the budget project of ICiG SB RAS No. FWNR-2022-0020.

Список литературы/References

1. Northoff G., Heinzel A., de Greck M. et al. Self-referential processing in our brain--a meta-analysis of imaging studies on the self. *Neuroimage*. 2006;31(1):440-57. doi 10.1016/j.neuroimage.2005.12.002
2. Knyazev G.G., Savostyanov A.N., Bocharov A.V., Kuznetsova V.B. Depressive symptoms and autobiographical memory: A pilot electroencephalography (EEG) study. *J Clin Exp Neuropsychol*. 2017;39(3):242-256. doi 10.1080/13803395.2016.1219318
3. Savostyanov A., Tamozhnikov S., Bocharov A. et al. The effect of meditation on comprehension of statements about one-self and others: A pilot ERP and behavioral study. *Front Hum Neurosci*. 2020;13:437. doi 10.3389/fnhum.2019.00437

doi 10.18699/bgrs2024-10-06

ЭЭГ реакции у здоровых детей и детей с нарушениями поведенческого контроля в условиях исполнения заданий в стоп-сигнал парадигме

Зорина К.А.[1]*, Дашкевич Г.Э.[2], Савостьянов А.Н.[1, 3, 4]

[1] *Институт медицины и психологии им. В. Зельмана Новосибирского государственного университета, Новосибирск, Россия*

[2] *Частный нейропсихологический центр, Новосибирск, Россия*

[3] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[4] *Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия*

* *k.zorina1@g.nsu.ru*

Ключевые слова: исполнительный контроль; стоп-сигнал парадигма; ЭЭГ; вызванные потенциалы головного мозга

Мотивация и цель: Исполнительный контроль над поведением – совокупность нейрофизиологических процессов, обеспечивающих возможность целенаправленного управления собственными двигательными реакциями в изменяющихся условиях внешней среды [1]. Исполнительный контроль включает в себя два компонента: активационный контроль, направленный на достижение поставленных целей, и тормозный контроль – подавление неадекватных действий. Парадигма стоп-сигнал (ССП) – экспериментальный метод, позволяющий оценить индивидуальные особенности активационного и тормозного контроля над поведением. В ходе ССП экспериментов участник должен либо быстро реагировать на один из нескольких стимулов нажатием кнопки, либо подавлять уже начатое действие, если после целевого стимула появляется стоп-сигнал [2]. В работах Савостьянова с соавторами [3] были выявлены нейрофизиологические маркеры, характеризующие мозговые механизмы активационного и тормозного контроля у здоровых взрослых. ССП широко применяется для исследования нейрофизиологических причин нарушений исполнительного контроля в разных группах пациентов [4]. Чаще всего при сравнении групп пациентов и здоровых контрольных участников в качестве показателей мозговой активности используются средние для каждого участника показатели амплитуды и латентности вызванных потенциалов, возникающих при выполнении активационных и тормозных заданий ССП. Однако можно предположить, что в качестве маркера расстройства могут выступать не средние значения ERP, а показатели внутрииндивидуальной или межиндивидуальной вариабельности мозговых ответов. Мы предполагаем, что вариабельность амплитуды и латентности ERP у детей с нарушениями поведенческого контроля будет существенно выше, чем в группе здоровых детей. Мы планируем сравнить показатели вариабельности амплитуды и латентности ERP пиков в условиях выполнения заданий парадигмы стоп-сигнал у здоровых детей и детей с нарушениями поведенческого контроля.

Методы и алгоритмы: В качестве экспериментальной группы были обследованы 20 школьников в возрасте 7–13 лет (12 мальчиков и 8 девочек), которые проходили

коррекцию в частном нейропсихологическом центре в связи с наличием у них нарушений поведенческого контроля. У этих детей наблюдалось снижение успеваемости в школе, нарушение внимания, слухоречевой памяти, праксиса, сформированности «образа тела» и пространственных представлений. В качестве контрольной группы были приглашены 20 детей того же возраста, у которых не было отмечено отклонений в поведенческом контроле. ЭЭГ обследование проводилось на основе парадигмы стоп-сигнал, модифицированной А.Н. Савостьяновым и его коллегами [3]. Испытуемым предлагалось быстро нажимать на одну из двух кнопок после появления одного из двух целевых сигналов либо останавливать движение после стоп-сигнала. Всего участники выполняли 95 испытаний на активационный контроль и 35 заданий на тормозный контроль, предъявляемых в случайном порядке. ЭЭГ регистрировалась при выполнении заданий при помощи 64-канального усилителя NVX, Россия.

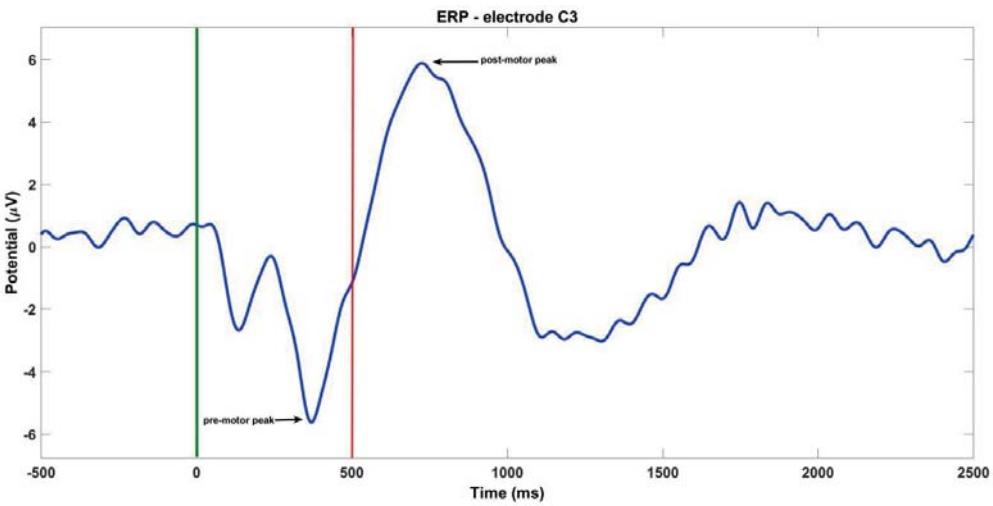


Рис. 1. Амплитудно-временной график событийных потенциалов в условиях активации SSP. Зеленая линия указывает момент появления целевого стимула, красная линия — среднее время нажатия кнопки

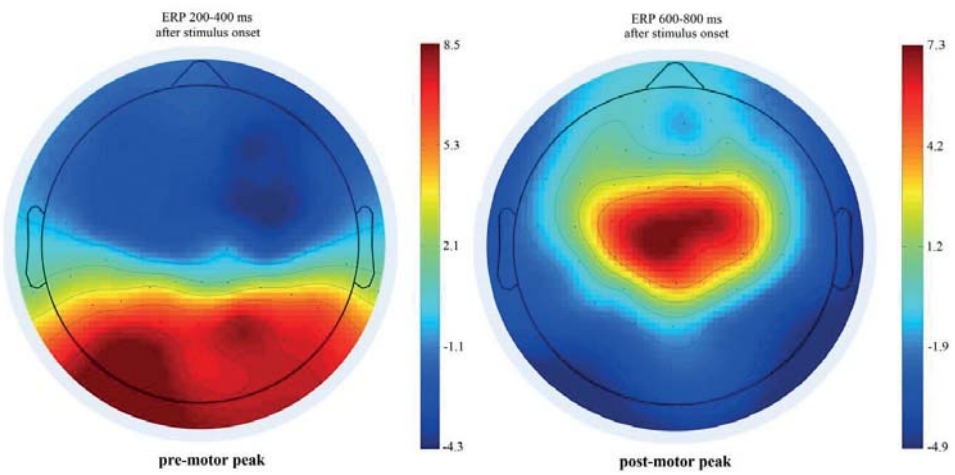


Рис. 2. Корковое распределение премоторных (слева) и постмоторных (справа) пиков

Артефакты подавлялись при помощи анализа независимых компонент. В качестве нейрофизиологических маркеров для условия активационного контроля вычислялась вариабельность амплитуды и латентности двух ERP пиков – премоторного компонента во временном окне 200–400 мс после появления целевого стимула и постмоторного компонента во временном окне 600–800 мс после целевого стимула (рис. 1 и 2). Для условия тормозного контроля оценивалась вариабельность амплитуды и латентности во временном окне 200–300 мс после появления стоп-сигнала.
Результаты: По предварительным данным для детей с нарушениями поведенческого контроля характерно увеличение как внутрииндивидуальной, так и внутригрупповой вариабельности для всех выбранных ERP пиков. При этом средняя амплитуда и латентность пиков в контрольной и экспериментальной группе достоверно не различаются ни для активационного, ни для тормозного условий в парадигме стоп-сигнал.
Выводы: Вариабельность амплитуды пиков ERP в парадигме стоп-сигнал может использоваться в качестве нейрофизиологического маркера нарушений в самоконтроле поведения.
Финансирование: Исследование поддержано за счет средств бюджетного проекта ИЦиГ СО РАН № FWNR-2022-0020.

EEG reactions in healthy children and children with behavioral control disorders during performing tasks in the stop-signal paradigm

Zorina K.A.[1]*, Dashkevich G.E.[2], Savostyanov A.N.[1, 3, 4]

[1] *V. Zelman Institute for Medicine and Psychology, Novosibirsk State University, Novosibirsk, Russia*

[2] *Private Neuropsychology Center, Novosibirsk, Russia*

[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[4] *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

* *k.zorina1@g.nsu.ru*

Key words: executive control; stop-signal paradigm; EEG; event-related potentials

Motivation and Aim: Executive behavioral control is a complex of neurophysiological processes that provide the ability to the self-control of behavior in a changing environment. [1]. Executive control includes two components – (1) activation control, aimed at achieving goals, and (2) inhibitory control – suppression of no-longer required or inappropriate actions. The stop-signal paradigm (SSP) is an experimental method that allows assessing the individual characteristics of activation and inhibitory control over behavior. During ERP experiments, the participant has to either quickly respond to one of several target stimuli by pressing a button, or suppress an action that has already begun if the stop-signal appears after the target stimulus [2]. The study of Savostyanov et al. [3] identified neurophysiological markers characterizing the brain mechanisms of activation and inhibitory control in healthy adults. Event-related potentials (ERP) is widely used to study the neurophysiological causes of executive control disorders in different groups of patients [4]. Most often, when comparing groups of patients and healthy control participants, the average amplitude and latency of ERPs arising during

the performance of activation and inhibitory SSP tasks for each participant are used as indicators of brain activity. However, it can be assumed that it is not the average ERP values that may serve as a marker of the disorder, but rather the indicators of intra-individual or inter-individual variability of brain responses. We hypothesize that the variability of ERP amplitude and latency in children with behavioral control disorders will be significantly higher than in the group of healthy children. We plan to compare the variability of the amplitude and latency of ERP peaks during the performance of stop-signal tasks in healthy children and children with behavioral control disorders.

Methods and Algorithms: As an experimental group, 20 school children aged 7–13 years (12 boys and 8 girls) were examined, who underwent correction in a private neuro-psychological center due to the presence of behavioral control deviations. These children experienced a decline in school performance, impairment of attention, auditory-verbal memory, praxis, formation of "body image" and spatial concepts. 20 children of this age who had no deviations in behavioral control were invited as a control group. The EEG examination was carried out on the basis of the stop-signal paradigm, modified by A.N. Savostyanov et al. [3]. Subjects were asked to quickly press one of two buttons after the appearance of one of two target signals, or to stop moving after a stop-signal. In total, participants completed 95 activation control tasks and 35 inhibitory control tasks, presented in random order. The EEG was recorded while performing tasks using a 64-channel amplifier NVX, Russia. Artifacts were suppressed using independent component analysis. As neurophysiological markers for the activation control condition, we calculated the variability of the amplitude and latency of two ERP peaks – (1) the premotor component in the time window 200–400 ms after the appearance of the target stimulus, and (2) the postmotor component in the time window 600–800 ms after the target stimulus (Fig. 1 and 2). For the inhibitory control condition, amplitude and latency variability were assessed in a time window of 200–300 ms after the onset of the stop-signal.

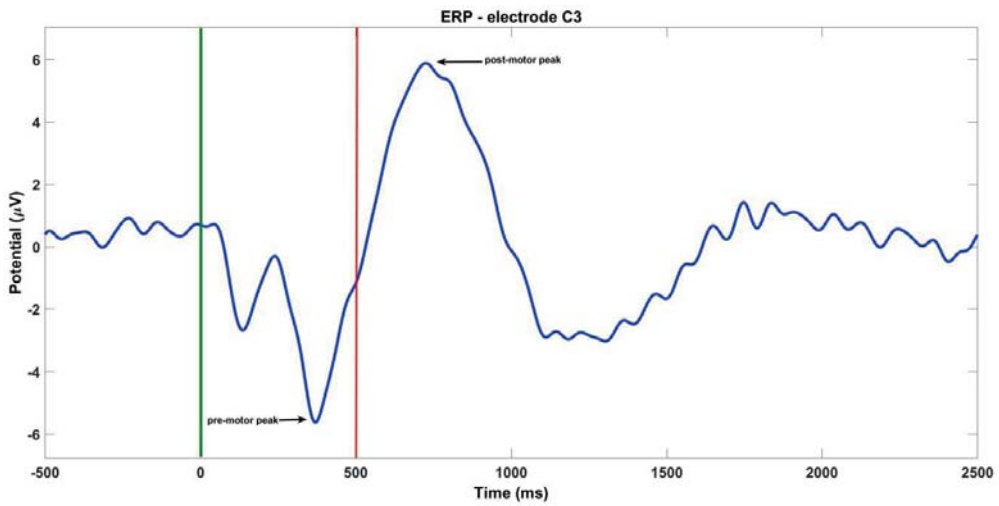


Fig. 1. The amplitude-time plot of event-related potentials under activation condition of SSP. The green line indicates a moment of target-stimulus onset, the red line indicates averaged time of a button pressing

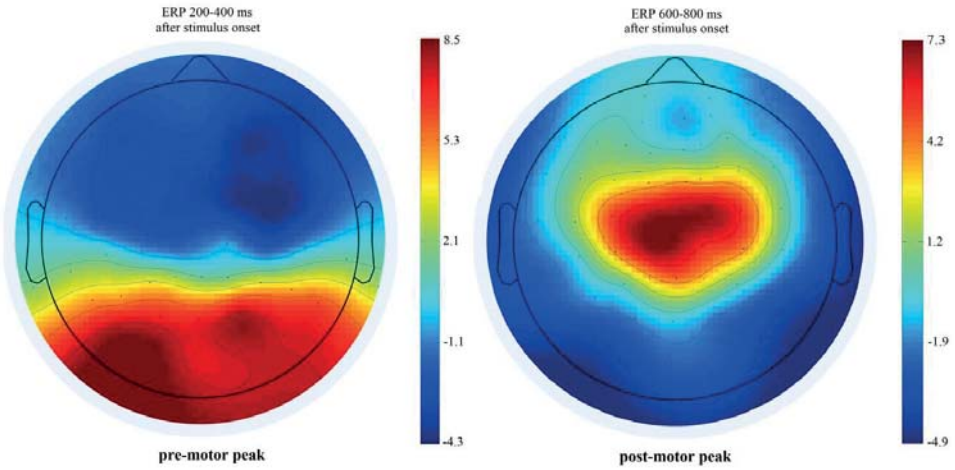


Fig. 2. The cortical distribution of pre-motor (left) and post-motor (right) ERP peaks

Results: According to preliminary data, children with behavioral control disorders are characterized by an increase in intra-individual and intra-group variability for all selected ERP peaks. At the same time, the average amplitude and latency of peaks in the control and experimental groups do not differ significantly for either activation or inhibitory conditions in the stop signal paradigm.
Conclusion: Amplitude and latency variability of ERP peaks in the stop-signal paradigm can be used as a neurophysiological marker of behavioral control disorder.
Funding: The study is supported by the budget project of the Institute of Cytology and Genetics SB RAS No. FWNR-2022-0020.

Список литературы/References

1. Logan G.D., Cowan W.B., Davis K.A. On the ability to inhibit simple and choice reaction time responses: A model and a method. *J Exp Psychol Hum Percept Perform*. 1984;10:276
2. Band G.P.H., van der Molen M.W., Logan G.D. Horse-race model simulations of the stop-signal procedure. *Acta Psychol*. 2003;112(2):105-142. doi 10.1016/s0001-6918(02)00079-3
3. Savostyanov A.N., Tsai A.C., Liou M. et al. EEG-correlates of trait anxiety in the stop-signal paradigm. *Neurosci Lett*. 2009;449(2):112-116. doi 10.1016/j.neulet.2008.10.084
4. Kaiser A., Aggensteiner P.M., Baumeister S., Holz N.E., Banaschewski T., Brandeis D. Earlier versus later cognitive event-related potentials (ERPs) in attention-deficit/hyperactivity disorder (ADHD): A meta-analysis. *Neurosci Biobehav Rev*. 2020;112:117-134. doi 10.1016/j.neubiorev.2020.01.019

doi 10.18699/bgrs2024-10-07

Онтогенез социальной мотивации, адресованной человеку, у доместицированных лисиц

Мухамедшина И.[1]*, Бесогонова К.[2], Румак А.[2], Харламова А.[1]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Пермский государственный университет, Пермь, Россия*

* *aden_66@mail.ru*

Ключевые слова: доместикация; лисицы; аффилиативное поведение; социальная мотивация; онтогенез поведения

Мотивация и цель: Уже на ранних этапах эксперимента по доместикации лисиц было показано, что отбор на снижение оборонительной реакции по отношению к человеку и усиление доместикационного поведения пролонгируют чувствительный период ранней социализации [1, 2]. Это выражается в том, что на протяжении онтогенеза у ручных лисиц исследовательская активность преобладает над неофобией, в то время как у неселекционируемых лисят в 1.5 мес. возникает реакция страха, которая блокирует исследование животным незнакомой обстановки. Мы продолжаем изучать изменения поведения и социальной мотивации, адресованной человеку, у лисиц из доместицируемой популяции в процессе их онтогенеза на современном этапе отбора.

Методы и алгоритмы: Лисята-самцы из доместицируемой популяции ИЦиГ СО РАН были протестированы нами дважды: впервые в возрасте 3.5–4 мес., затем повторно в 6.5–7 мес. Общее число составило 20 животных. Лисята содержались в стандартных условиях зверофермы, в индивидуальных клетках 90×90 см. Для проведения тестов лисят в возрасте 3.5–4 мес. без предварительной адаптации помещали в незнакомый для них вольер 6×6 м, разделенный на две равные части перегородкой, составленной из различных материалов (ширма, листы ДВП, столешница и пр.). При изготовлении перегородки в ней преднамеренно было оставлено несколько мест для возможного перехода лисицы с одной стороны вольера на другую. Животным давали 2 мин на исследование вольера, затем начинали ласково разговаривать с ними, побуждая их подойти к незнакомому человеку. Тестирование проводили три исследователя. Если лисенок подходил к людям, экспериментаторы призывали его следовать за собой, переходя с одной стороны перегородки на другую и обратно (человек имел возможность перешагнуть через преграду в разных местах; лисица самостоятельно искала варианты для перехода). В разделе «результаты» представлены данные, измеренные в течение первой минуты, с момента, когда экспериментаторы начинали перешагивать через преграду и звать лисенка к себе.

Если лисенок следовал за людьми, преодолевая преграду, переходили к следующему этапу тестирования. Животное знакомили со звенящей игрушкой (размером с мышь), которую экспериментатор шевелил при помощи веревки, имитируя убегающую добычу. В течение этого этапа (1 мин) лисенка побуждали следовать за игрушкой через преграду. После знакомства с игрушкой один из экспериментаторов оставался по одну сторону от преграды, он полностью закрывался белым халатом, выставляя наружу только руку со звенящей игрушкой.

Двое других экспериментаторов уходили на обратную сторону преграды, садились напротив и в течение 1 мин активно и эмоционально привлекали внимание лисенка, побуждая его подойти к ним. В возрасте 7 мес. проводили повторное тестирование в тех же условиях и по тому же алгоритму с лисицами, которые в возрасте 3.5–4 мес. успешно завершили все этапы эксперимента.

Результаты: Преобладание боязни новой обстановки над исследовательским поведением наблюдалось только у 5 лисят, остальные 15 животных успешно прошли тестирование. Ниже приведены данные по лисятам, успешно завершившим все этапы эксперимента. Лисята из доместицируемой популяции следовали за человеком благодаря очень сильной социальной мотивации, никакой подкормки животным не давали, не заманивали игрушкой. Лисята воспринимали человека как социального партнера, что облегчало их адаптацию к новым условиям. В данном эксперименте мы измеряли число переходов ручных лисиц за экспериментатором через преграду в течение 1 мин (среднее число составило 4.33±0.33; n = 15). Животные находили разнообразные способы преодоления преграды в процессе следования за человеком. Все перечисленные варианты переходов через преграду проиллюстрированы на видеозаписи: https://rutube.ru/video/c51261ecc46a0b71d36d23c83539d5dd/?r=plwd. Среднее число использованных лисицами вариантов составило 3.46±0.24; n = 15), при этом общее количество переходов за 1 мин положительно коррелировало с количеством вариантов (R = +0.67; P≤0.05, критерий Спирмена r-Spearman's). Поведение животных в течение теста характеризуют следующие показатели: 5.46±1.88 – среднее время (с), в течение которого лисята обнюхивали пол или стены в вольере; 23±2.34 – среднее время (с), в течение которого лисята находились в радиусе 1 м около человека (в движении и в статике); 11.2±2.2 – латентное время (с) до первого прохода лисенка, следующего за человеком, через преграду.

В возрасте 6.5–7 мес. у части исследованных нами лисиц снизилась мотивация подходить к человеку и следовать за ним. Один из лисят, в более раннем возрасте успешно проходивший данный тест, ни разу не прошел за человеком через преграду. Однако большинство лисят продолжали следовать за экспериментатором, преодолевая преграду. Среднее число переходов составило 4.2±0.6; n = 15; среднее число использованных лисицами способов преодоления преграды – 3.13±0.4; n = 15. Общее количество переходов за 1 мин, как и в возрасте 3.5–4 мес., положительно коррелировало с количеством вариантов (R = +0.85; P≤0.05, критерий Спирмена r-Spearman's). Статистически значимых различий по количеству переходов между двумя возрастными группами показано не было.

По всем остальным показателям, характеризующим мотивацию лисиц следовать за человеком, нами были получены значимые различия. Среднее время (с), в течение которого лисята обнюхивали пол или стены в вольере у сеголеток (6.5–7 мес.) увеличилось и составило 12.6±3 (p≤0.01, Wilcoxon Matched Pairs Test, по сравнению с этим показателем в возрасте 3.5–4 мес.). Среднее время (с), в течение которого лисицы находились в радиусе 1 м около человека (в движении) либо интересовались стоящим человеком (смотрит в лицо человека, принюхивается) у сеголеток снизилось до 14.3±2.1 (p≤0.01, Wilcoxon Matched Pairs Test, по сравнению этим показателем в возрасте 3.5–4 мес.). Латентное время (с) до первого прохода лисенка через преграду у сеголеток увеличилось до 32.3±11

(p≤0.05, Wilcoxon Matched Pairs Test, по сравнению с этим показателем в возрасте 3.5–4 мес.).
На следующем этапе тестирования один из экспериментаторов пытался отвлечь лисенка игрушкой, а двое других звали животное к себе. В течение 1 мин тестирования лисята в возрасте 3.5–4 мес. достоверно больше времени (с) проводили с человеком, чем сеголетки: 13.8±5 по сравнению с 5.8±3 (p≤0.05, Wilcoxon Matched Pairs Test). По времени, в течение которого лисята играли с игрушкой на данном этапе теста, различий между двумя возрастными периодами нами не получено.
Выводы: 1. Чем выше у доместицируемых лисиц социальная мотивация, адресованная человеку, тем больше раз они преодолевают преграду за минуту теста. Число переходов через преграду положительно коррелирует с числом разных способов ее преодоления.
2. У доместицируемых лисиц наблюдается снижение социальной мотивации, адресованной человеку, в возрасте 6.5–7 мес. по сравнению с возрастом 3.5–4 мес.
Финансирование: Исследование поддержано грантом РНФ № 21-44-04405.

Ontogenesis of social motivation addressed to human in domesticated foxes

Mukhamedshina I.[1]*, Besogonova K.[2], Rumak A.[2], Kharlamova A.[1]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Perm State University, Perm, Russia*
* *aden_66@mail.ru*

Key words: domestication; foxes; social motivation; behavior ontogenesis; affiliative behavior

Motivation and Aim: Selection of foxes on decreasing in the defensive reaction towards human and an increasing domestication behavior prolongs the sensitive period of early socialization, as was demonstrated at the early stages of the experiment [1, 2]: exploratory activity of tame foxes dominates over neophobia during ontogenesis, while in unselected fox cubs at 1.5 months a fear reaction occurs, which stops the animal's exploration in new environment. We continue to study changes in behavior and social motivation addressed to humans in the ontogenesis of domesticated foxes at the current stage of selection.
Methods and Algorithms: Male fox cubs from the domesticated population of the Institute of Cytology and Genetics SB RAS were tested two times: at the age of 3.5–4 months, as well as at the age of 6.5–7 months. The total number was 20 animals. The fox cubs were kept in standard fur farm conditions, in individual cages 90×90 cm. For first testing, fox cubs were placed in an unfamiliar facility of 6×6 m, without any preliminary adaptation. The enclosure divided into two equal parts by obstacle made of various materials (screen, fiberboard sheets, table top, etc.). When making the obstacle, several places were intentionally left in it for the fox to move from one side of the enclosure to other. Each animals were given 2 minutes to explore the enclosure, then experimenters began to talk to them affectionately, encouraging them to approach the stranger. Testing was carried out by three researchers. If the fox cub approached people, the experimenters called on him to follow them, crossing from one side of the obstacle

to the other and back (the person had the opportunity to step over the barrier in different places; the fox independently looked for options for crossing). This stage of the test lasted 1 minute.

If the fox cub followed the people, overcoming the obstacle, we began next stage of testing. The animal was introduced to a ringing toy (the size of a mouse), which the experimenter moved with a rope, imitating fleeing prey. During this stage (1 min), the fox cub was encouraged to follow the toy through the obstacle. After getting acquainted with the toy, one of the experimenters remained on one side of the barrier; he was completely covered with a white robe, exposing only his hand with the ringing toy. Two other experimenters went to the opposite side of the obstacle, sat opposite and for 1 minute actively and emotionally attracted the attention of the fox cub, encouraging him to approach them. At the age of 7 months we repeated testing. The experiment was carried out under the same conditions and according to the same algorithm with foxes that were 3.5–4 months old successfully completed all stages of the experiment.

Results: Only in 5 fox cubs neofobia was predominant on fear in new enclosure, the remaining 15 animals successfully passed the test. We will present the results only on fox cubs that successfully completed all stages of the experiment. Domesticated foxes followed people by very strong social motivation – during experiment this behavior was not rewarded by food or toy. The fox cubs perceived human as a social partner, which facilitated their adaptation to new environment. In this experiment we measured the number of transitions through the barrier in tame foxes within 1 minute (the average number was 4.33±0.33; n=15). Animals found various ways to overcome obstacles while following experimenter. All of the listed options for crossing an obstacle are illustrated in the video: https://rutube.ru/video/c51261ecc46a0b71d36d23c83539d5dd/?r=plwd.

The average number of variants used by foxes was 3.46±0.24; n=15). The total number of transitions in 1 minute was positively correlated with the number of variants (R=+0.67; P≤0.05, Spearman's r-Spearman's test). The behavior of the animals during the test is characterized by the following indicators: 5.46±1.88 – average time (sec) during which the fox cubs sniffed the floor or walls in the enclosure; 23±2.34 – average time (sec) during which the fox cubs were within a radius of 1 m near a person (in motion and static); 11.2±2.2 – latent time (sec) before the first passage of a fox cub following a person through an obstacle.

At the age of 6.5–7 months some of the foxes we studied had decreased motivation to approach and follow people. One of the foxes, who successfully passed this test at an earlier age, not followed people through an obstacle. However, most of the fox cubs continued to follow the experimenter, overcoming the obstacle. The average number of transitions was 4.2±0.6; n=15; the average number of variants used by foxes to overcome the obstacle was 3.13±0.4; n=15). The total number of transitions in 1 minute, as at the age of 3.5–4 months, was positively correlated with the number of variants (R=+0.85; P≤0.05, Spearman's r-Spearman's test). There were no statistically significant differences in the number of transitions between the two ages.

We obtained significant differences for all other parameters characterizing the motivation of foxes to follow people. The average time (sec) during which fox cubs sniffed the floor or walls of enclosure in adolescence (6.5–7 months) increased to 12.6±3 (p≤0.01, Wilcoxon Matched Pairs Test, compared with same parameter at the age of 3.5–4 months). The average time (sec) during which foxes were within a circle of 1 m near a experimenter (in motion), or were interested in a standing experimenter (looks at the people's face, sniffs) among in adolescence decreased to 14.3±2.1 (p≤0.01, Wilcoxon

Matched Pairs Test, compared with this parameter at the age of 3.5–4 months). The latent time (sec) before the first passage of a fox cub through an obstacle in adolescence increased to 32.3±11 ($p \leq 0.05$, Wilcoxon Matched Pairs Test, compared with this parameter at the age of 3.5–4 months).

At the next stage of testing, one of the experimenters tried to distract the fox cub with a toy, while the other two called the animal to them. During 1 minute of testing, fox cubs aged 3.5–4 months, spent significantly more time (sec) with humans than in adolescence: 13.8±5 compared to 5.8±3 ($p \leq 0.05$, Wilcoxon Matched Pairs Test). In terms of the time during which the fox cubs played with the toy during this stage of the test, we not found differences between the two age periods.

Conclusion: 1. Domesticated foxes whose social motivation addressed to human is higher, overcome the obstacle more times per minute of the test. The number of crossings over an obstacle is positively correlated with the numbers of different variants to overcome it.

2. Social motivation addressed to humans in domesticated foxes is a decrease at the age of 6.5–7 months, compared to the age of 3.5–4 months.

Funding: The study is supported by RSF grant No. 21-44-04405.

Список литературы/References

1. Беляев Д.К., Плюснина И.З., Трут Л.Н. Физиологические границы чувствительного периода первичной социализации у серебристо-черных лисиц, их изменение в процессе доместикации. *Журнал эволюционной биохимии и физиологии*. 1986;22(6):555-562
2. Belyaev D.K., Plyusnina I.Z., Trut L.N. Domestication in the silver fox (Vulpes fulvus Desm): Changes in physiological boundaries of the sensitive period of primary socialization. *Appl Anim Behav Sci*. 1985;13(4):359-370

doi 10.18699/bgrs2024-10-08

Влияние систем приближения/избегание на особенности активации мозга при восприятии природных стимулов в формате 2D и 3D

Разумникова О.*, Давыдов А., Бакаев М.
Новосибирский государственный технический университет, Новосибирск, Россия
** razoum@mail.ru*

Ключевые слова: системы мотивации; ЭЭГ; альфа ритм; природная среда; виртуальная реальность

Обоснование и цель: Появляется все больше доказательств воздействия природной среды (green space), в том числе представленной в виртуальной реальности (VR), как инструмента улучшения эмоциональной регуляции, нейрореабилитации, стимуляции положительных эмоций и ослабления стресса [1–3]. Показано, что для положительного эффекта природной среды достаточно кратковременного десятиминутного пребывания в ней [4], причем восстановительное воздействие green space, предъявленного в формате 2D или 3D, не отличается в условиях анализа экспериментально созданного стресса [5]. Отмечено, однако, существенное разнообразие показателей индивидуальных впечатлений и психофизиологического состояния участников исследований [6–9], вследствие чего возникают вопросы о выяснении закономерностей природного воздействия на эмоциональное состояние человека или причин отсутствия такого эффекта.
Одним из таких факторов может быть влияние функций систем избегания / приближения (BIS/BAS), связанных с выраженностью полушарной асимметрии в активации префронтальных областей коры головного мозга: BIS – с правополушарной, а BAS -левополушарной [10]. Несмотря на давно возникший интерес к роли этих систем, влияющих на психическое состояние вследствие доминирующей мотивации в поведении, единого мнения относительно их значения пока не сложилось, так как известна гипотеза о том, что функции BAS могут быть представлены билатеральной активностью лобной коры [11]. Поэтому целью настоящего исследования стало выяснение связи самооценки показателей BIS/BAS и частотно-пространственной организации активности мозга при восприятии природной среды в формате 2D и 3D.
Материалы и методы: В исследовании принимали участие 20 студентов университета (18-23 лет, половину их них составили мужчины). Для просмотра изображений в формате 3D использовали шлем виртуальной реальности Oculus Rift S с фирменным программным обеспечением. В трех экспериментальных ситуациях: фон с открытыми глазами (ФОГ) и фильм природного содержания («Прогулка по лесу», рис. 1), просматриваемый в формате 2D и 3D, в течение 3–5 минут регистрировали ЭЭГ (62 канала, Neuvo SynAmps2 System). Удаление артефактов и расчет мощности ритмов с использованием преобразования Фурье выполняли с применением EEGLAB. Так как доминирование систем BIS или BAS связывают с асимметрией активности префронтальных областей коры согласно соотношению мощности альфа ритма [10], мы представляем результаты,

полученные при анализе биопотенциалов этой частоты. Для оценки показателей систем BIS/BAS использовали методику Грея–Уилсона [12].

Рис. 1. Пример кадра из фильма, предъявляемого в формате 2D и 3D

Результаты: При сравнении показателей BIS/BAS обнаружено преобладание системы BAS (11.2±3.3) по сравнению с BIS (8.2±4.8) ($p < 0.04$) с доминированием BAS у 60% участников исследования. Корреляционный анализ показателей этих систем и мощности альфа1 ритмов выявил значимые связи (согласно критерию Спирмена) для ситуаций ФОГ и 3D, которые представлены в табл. 1. При этом рассматриваются только корреляции, регионарно группирующиеся в рядом расположенных отведениях ($0.01 < p < 0.10$). В альфа1 диапазоне для ситуации 2D обнаружены положительные связи BAS и биопотенциалов только в отведениях Cz и C6 ($0.37 < Rs < 0.46$, $p < 0.10$). На альфа 2 частоте выявленные множественные связи показателей BAS представлены на рис. 2.

Таблица 1.

ФОГ ВЕО	Отв	F1	Fz	F2	F6	F12	FC2	FC6	C4	CP5	POz	PO4
	Rs	0.37	0.38	0.45	0.42	0.39	0.39	0.39	0.37	0.39	0.39	0.40
3D	Отв	F5	F7	F11	C2	Cz	C4	CPz				
	Rs	0.37	0.43	0.36	0.48	0.48	0.37	0.45				

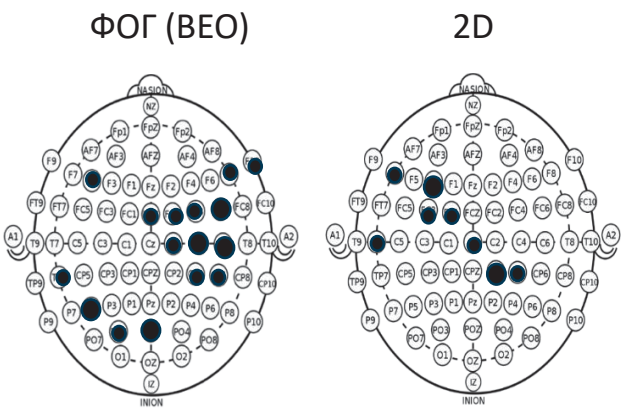


Рис. 2. Регионарные особенности связи показателей системы приближения и мощности альфа 2 ритма в фоне (ФОГ) и при предъявлении природных стимулов в формате 2D

В ситуации ФОГ они преимущественно охватывают фронтально-центральные области правого полушария и париетально-окципитальные левого, а в ситуации 2D – передние области левого, для ситуации 3D обнаружена единственная значимая связь для альфа 2 в отведении F7.
Выводы: Полученные результаты указывают на «преднастройку» активационного состояния мозга в соответствии с выраженностью мотивационной системы BAS, представленной билатерально, как это предложено ранее [11]. Отмеченная положительная связь BAS и мощности альфа осцилляций свидетельствует, что доминирование системы приближения дополнительно способствует эффекту релаксации, показанному при погружении в природную среду [1–5].

Impact of approach/avoidance systems on specificity of brain activation when perceiving natural stimuli in 2D and 3D format

Razumnikova O.*, Davidov A., Bakaev M.
Novosibirsk State Technical University, Novosibirsk, Russia
** razoum@mail.ru*

Key words: motivation systems; EEG; alpha rhythm; natural environment; virtual reality

Background and Aim: Recently, there has been more and more evidence of the impact of the natural environment (green space), including that presented in virtual reality (VR), as a tool for improving emotional regulation, neurorehabilitation, stimulating positive emotions and reducing stress [1–3]. It has been shown that positive effect of the natural environment may be obtained by a short ten-minute stay in it [4], and the restorative effect of green space, presented in 2D or 3D format, does not differ in the conditions of analysis of experimentally induced stress [5]. However, significant variability has been noted in the indicators of individual impressions and the psychophysiological state of research participants [6–9]. Thus, the studies are needed to determine the patterns of nature influence on a person's emotional state or the reasons for the absence of such an effect. One of these factors may be the impact of the functions of the avoidance/approach systems (BIS/BAS), reflected by the hemispheric asymmetry in activation of prefrontal brain areas. Namely, BIS is associated with the right hemispheric activation whereas BAS is presented by the left hemispheric activation [10]. Despite the long-standing interest in the role of these systems, influencing the mental state due to the dominant motivation in behavior, there is no consensus on their significance yet, since there is a known hypothesis that bilateral frontal cortical activity may be related to the BAS [11]. Therefore, the aim of this study was to determine the relationships between self-assessment of BIS/BAS indicators and the frequency-spatial organization of brain activity during the perception of the natural environment in 2D and 3D format.
Materials and Methods: The study involved 20 university students (18-23 years old, half of them were men). To view images in 3D format, we used an Oculus Rift S virtual reality helmet with proprietary software. EEG was recorded for 3–5 minutes (62 channels, Neuvo SynAmps2 System) in three experimental situations: a background with eyes open (BEO) and a film with natural content (A walk in the forest, Fig. 1), viewed in 2D and 3D format. A removal of artifacts and calculation of rhythm power by Fourier transform were performed using EEGLAB. Since the dominance of the BIS or BAS systems is associated with asymmetry of activation in the prefrontal cortex

according to the alpha power index [10], we present the results obtained from the analysis of biopotentials of this frequency. To measure the indices of BIS/BAS systems, Gray–Wilson Personality Inventory Short Form was used [12].

Fig. 1. An example of a frame from a film presented in 2D and 3D formats

Results: When comparing BIS/BAS indices, a predominance of the BAS system (11.2±3.3) was found compared to BIS (8.2±4.8) ($p < 0.04$) with BAS dominance in 60% of study participants. Correlation analysis of the indices of these systems and the power of alpha1 rhythms revealed significant relationships (according to the Spearman criterion) for the BEO and 3D situations, which are presented in the Table 1. Only correlations that are regionally grouped in nearby electrode sites are considered ($0.01< p <0.10$). In the alpha1 range for the 2D situation, positive relations between BAS and biopotentials were found only in the Cz and C6 sites ($0.37 < Rs < 0.46$, $p < 0.10$). At the alpha 2 frequency, the identified multiple relationships of BAS indicators are presented in Fig. 2.

Table 1.

ФОГ	Отв	F1	Fz	F2	F6	F12	FC2	FC6	C4	CP5	POz	PO4
BEO	Rs	0.37	0.38	0.45	0.42	0.39	0.39	0.39	0.37	0.39	0.39	0.40
3D	Отв	F5	F7	F11	C2	Cz	C4	CPz				
	Rs	0.37	0.43	0.36	0.48	0.48	0.37	0.45				

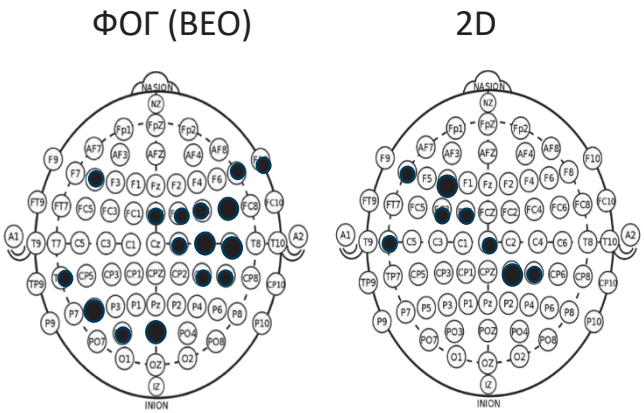


Fig. 2. Patterns of regional relationships between indices of the BAS and the alpha 2 power in the background (BEO) and while presentation of natural stimuli in 2D format

In the BEO situation, the pattern of correlations is presented predominantly in the frontal-central areas of the right hemisphere and the parietal-occipital areas of the left one, and in the 2D situation – in the anterior areas of the left hemisphere; for the 3D situation, the only significant relation was found for alpha 2 in the F7 site.

Conclusion: The results obtained indicate a "pre-tuning" of the activation state of the brain by impact of the BAS motivational system, represented bilaterally, as previously proposed [11]. The noted positive relationship between BAS and the power of alpha oscillations indicates that the dominance of the approach system additionally contributes to the relaxation effect shown during immersion in the natural environment [1–5].

Список литературы/References

1 Colombo D., Díaz-García A., Fernandez-Álvarez J., Botella C. Virtual reality for the enhancement of emotion regulation. *Clin Psychol Psychother*. 2021;28(3):519-537. doi 10.1002/cpp.2618

2 Jung M., Apostolova L.G., Moser D.K., Gradus-Pizlo I., Gao S., Rogers J.L., Pressler S.J. Virtual reality cognitive intervention for heart failure: CORE study protocol. *Alzheimers Dement (NY).* 2022;8(1):e12230. doi 10.1002/trc2.12230

3 Liu P., Liu J., Fernandez J., Zou Q., Lin M. Positive affect and natural landscape in virtual reality: A systematic review comparing interventions, measures, and outcomes. *J Environ Psychol.* 2023;88:102011. doi 10.1016/j.jenvp.2023.102011

4 Meredith G.R., Rakow D.A., Eldermire E.R.B., Madsen C.G., Shelley S.P., Sachs N.A. Minimum time dose in nature to positively impact the mental health of college-aged students, and how to measure it: A soping review. *Front Psychol*. 2020;10:2942. doi 10.3389/fpsyg.2019.02942

5 Suseno B., Hastjarjo T.D. The effect of simulated natural environments in virtual reality and 2D video to reduce stress. *Front Psychol*. 2023;14:1016652. doi 10.3389/fpsyg.2023.1016652

6 Разумникова О.М., Варнавский И.Н. Особенности эмоционального отношения к природным и техногенным ландшафтам в молодом и пожилом возрасте. *Валеология*. 2017;4:55-61.
[Razumnikova O.M., Varnavslii I.N. Specificity of emotional reactions to natural and technogenic landscapes in young and old people. *Valeologiya.* 2017;4:55-61 (in Russian)]

7 Davidov A., Razumnikova O., Bakaev M. Nature in the heart and mind of the beholder: Psycho-emotional and EEG differences in perception of virtual nature due to gender. *Vision (Basel).* 2023;7(2):30. doi 10.3390/vision7020030

8 Montana J.I., Matamala-Gomez M., Maisto M., Mavrodiev P.A., Cavalera C.M., Diana B., Mantovani F., Realdon O. The benefits of emotion regulation interventions in virtual reality for the improvement of wellbeing in adults and older adults: A systematic review. *J Clin Med*. 2020;9(2):500. doi 10.3390/jcm9020500

9 Mygind L., Kjeldsted E., Hartmeyer R., Mygind E., Stevenson M.P., Quintana D.S., Bentsen P. Effects of public green space on acute psychophysiological stress response: a systematic review and meta-analysis of the experimental and quasi-experimental evidence. *Environ Behav*. 2019;53:184-226. doi 10.1177/0013916519873376

10 Sutton S.K., Davidson R.J. Prefrontal brain asymmetry: a biological substrate of the behavioral approach and inhibition systems. *Psychol Sci.* 1997;8:204-210

11 Harmon-Jones E. Clarifying the emotive functions of asymmetrical frontal cortical activity. *Psychophysiology.* 2003;40:838-848

12 Князев Г.Г., Слободская Е.Р., Сафронова М.В. Краткая форма личностного опросника Грея-Уильсона. *Вопросы психологии*. 2004;4:113-122
[Knyazev G.G., Slobodskaya E.P., Safronova M.V. Gray-Wilson Personality Inventory Short Form. *Voprosy Psikhologii*. 2004;4:113-122 (in Russian)]

doi 10.18699/bgrs2024-10-09

Взаимосвязь спектральных характеристик ЭЭГ покоя, индуцированных распознаванием своего и чужого лицевого видео, с показателями коллективизма у российских и китайских студентов

Савостьянов А.Н.[1, 2, 3]*, Сы Ция[3], Тянь Цзяхао[3], Сапрыгин А.Е.[1, 2], Кулешов Д.А.[2], Савостьянов В.А.[3], Бочаров А.В.[1]

[1] *Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[3] *Новосибирский государственный университет, Новосибирск, Россия*

* *a-sav@mail.ru*

Ключевые слова: ЭЭГ; условия функционального покоя; самореференция; коллективизм; межкультурные сравнения

Мотивация и цель: Активность головного мозга человека в условиях покоя вызывает большой интерес у современной нейрофизиологии. Концепция сетей покоя мозга основывается на данных фМРТ и ЭЭГ, показывающих наличие функциональных процессов, ассоциированных с размышлениями человека о самом себе [1]. Изучение таких процессов у здоровых людей открывает возможность для понимания нейрофизиологических механизмов самооценки и самореференции [2]. Кроме того, анализ мозговой активности в условиях покоя дает возможность выявить функциональные маркеры аффективных расстройств [3]. Остается открытым вопрос: насколько активность в условиях покоя зависит от событий, предшествовавших этим условиям? Можно предположить, что концентрация внимания испытуемых на внешних стимулах, отнесенных к самому себе или другим людям, вызовет изменения в активности покоя, регистрируемой некоторое время после такой концентрации.

Ранее нами было показано, что динамика мозговой активности в дефолт-системе мозга в существенной степени зависит от социокультурной принадлежности участников обследования [4]. При сравнении российских и тайваньских здоровых участников мы показали, что у тайванских участников процессы размышления о самом себе сильнее ассоциированы с передним (предлобным) хабом дефолт-системы, тогда как у российских участников эти процессы были сильнее связаны с задним (теменным) хабом системы. Мы предположили, что для представителей азиатской культуры более свойственны этические установки коллективизма, тогда как представители российской культуры более склонны к индивидуализму, что может служить одной из причин различий в мозговой активности покоя.

В нашем новом исследовании мы сопоставляем показатели спектральной плотности мощности ЭЭГ, записанной в интервалах между наблюдениями видеозаписей собственного лица участника, видеозаписи лица незнакомого ему человека и наблюдением пустого экрана компьютера. В качестве групп сравнения приглашались здоровые российские студенты и студенты из КНР, обучающиеся в России. Показатели коллективизма, оцениваемые при помощи психологических

опросников, а также возраст и пол испытуемых включены в анализ ЭЭГ данных в качестве дополнительных переменных.
Методы и алгоритмы:
Участники: 30 здоровых российских студентов и аспирантов (средний возраст 23.4 года, 13 мужчин, 17 женщин) и 30 китайских студентов и аспирантов, обучающихся в Новосибирском государственном университете (средний возраст 22.3 года, 16 мужчин, 14 женщин) участвовали в ЭЭГ обследовании. Все студенты заполняли комплект психологических опросников на коллективизм-индивидуализм, аффилиацию с семейными ценностями, степень фиксации внимания на самом себе, уровень личностной, ситуативной и социальной тревожности, склонность к депрессии. Во время эксперимента участники сидели в свето- и звукоизолированной комнате. ЭЭГ регистрировалась в три этапа: 1) 12 минут записи с повторяемыми двухминутными интервалами с закрытыми и открытыми глазами (всего 6 минут с закрытыми глазами и 6 минут с открытыми глазами), когда при открывании глаз испытуемый видел черный экран компьютера; 2) 12 минут записи с двухминутными чередованиями закрывания и открывания глаз, когда при открывании глаз испытуемый видел видеозапись собственного лица; 3) 12 минут записи с двухминутными чередованиями закрывания и открывания глаз, когда при открывании глаз испытуемый видел видеозапись лица незнакомого ему человека того же пола, что сам участник.
Регистрация и анализ ЭЭГ: электрическая активность мозга записывалась при помощи 128 ЭЭГ каналов + ЭОГ + ЭКГ через усилитель NVX-136, Россия, полоса пропускания 0.3–100 Гц, частота дискретизации сигнала 1000 Гц, референтный электрод Cz, заземление AFz. Глазодвигательные артефакты удалялись при помощи анализа независимых компонент. Для дальнейшего анализа использовали 100–120-секундные ЭЭГ записи при закрытых глазах в интервалах между наблюдением пустого экрана, своего или чужого лица. Плотность спектральной мощности в разных частотных диапазонах, ее зависимость от экспериментального условия, группы участников и их психологических характеристик анализировали при помощи программного пакета sLORETA.
Результаты: 1. Психологические сравнения выявили, что для китайских участников действительно характерен повышенный уровень показателей коллективизма в сравнении с российскими участниками по большинству использованных нами опросников. Однако оценка взаимодействия факторов пола и национальности показала, что более высокие показатели коллективизма характерны для китайцев мужского пола, чем для российских мужчин, тогда как китайские и российские женщины по этому показателю различались существенно слабее. Китайские и российские испытуемые не различались по степени склонности к депрессии и уровню личностной тревожности.
2. Сравнение трех экспериментальных условий в обеих группах участников выявило, что наблюдение своего лица в сравнении с наблюдением пустого экрана вызывает снижение спектральной плотности в лобной коре в диапазонах дельта-, альфа- и гамма-ритмов, тогда как наблюдение чужого лица в сравнении с наблюдением пустого экрана вызывало снижение спектральной плотности в лобной, теменной и латеральной височной коре в тех же диапазонах, что и наблюдение своего лица. В целом различия в спектральной плотности при разных экспериментальных условиях наблюдались в корковых структурах, входящих в состав дефолт-системы.

3. Межгрупповые различия между российскими и китайскими участниками были наиболее выражены в условиях покоя после наблюдения чужого лица и в меньшей степени – после наблюдения за собственным лицевым видео. У китайских участников наблюдалась большая спектральная плотность мощности альфа-ритма в лобных отделах коры.
4. Показатели коллективизма были ассоциированы с большей спектральной плотностью мощности в альфа-диапазоне в лобных отделах коры у обеих групп участников. Этот эффект был достоверен после наблюдения за своим лицом у российских участников и после наблюдения как за своим, так и за чужим лицом у китайских участников.
5. У всех женщин в сравнении с мужчинами, вне зависимости от их национальности, наблюдалась повышенная спектральная плотность мощности альфа-ритма в теменных и латеральных височных отделах коры после наблюдения за своим собственным лицом. Такие межполовые различия отсутствовали в условиях покоя после наблюдения пустого экрана или чужого лица.
Выводы: Нами предложена экспериментальная модель, позволяющая сравнивать электрическую активность головного мозга в условиях покоя, индуцированную предшествующей концентрацией внимания на стимулы, относящиеся к самому испытуемому, другому человеку или неодушевленному объекту. Эта модель была использована для изучения зависимости мозговой активности покоя от социокультурной принадлежности участников обследования. Анализ спектральной плотности в условиях покоя, следующих за наблюдением собственного лицевого видео, лицевого видео незнакомого человека или пустого экрана, выявил особенности функциональных состояний человека, зависящие от его пола, этнокультурной принадлежности и уровня коллективизма.
Финансирование: Исследование поддержано за счет средств гранта РНФ № 22-15-00142 «фМРТ и ЭЭГ корреляты фокуса внимания на собственной персоне как фактора предрасположенности к аффективным расстройствам». Работа А.Е. Сапрыгина и А.Н. Савостьянова по разработке методик анализа нейрофизиологических данных финансировалась из средств бюджетного проекта ИЦиГ СО РАН № FWNR-2022-0020.

Association of spectral characteristics of resting-state EEG induced by recognition of one's own and other's facial video with indexes of collectivism in Russian and Chinese students

Savostyanov A.N.[1, 2, 3]*, Si Qiya[3], Tian Jiahao[3], Saprygin A.E.[1, 2], Kuleshov D.A.[2], Savostyanov V.A.[3], Bocharov A.V.[1]

[1] *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *a-sav@mail.ru*

Key words: EEG; functional resting state; self-reference; collectivism; cross-cultural differences

Motivation and Aim: The activity of the human brain during resting state is of great interest to modern neuroscience. The concept of resting state networks is based on fMRI

and EEG data showing the presence of functional processes associated with a person's thoughts about oneself [1]. The study of such processes in healthy people opens up the possibility of understanding the neurophysiological mechanisms of self-esteem and self-reference [2]. In addition, analysis of brain activity during resting state makes it possible to identify functional markers of affective disorders [3]. The question remains open: to what extent does activity during resting state depends on the events that preceded this resting state? It can be assumed that concentration of the subjects' attention on external stimuli related to oneself or other people will lead to changes in resting activity recorded some time after said concentration. We have previously shown that the dynamics of brain activity in the default mode network largely depend on the socio-cultural affiliation of the survey participants [4]. When comparing Russian and Taiwanese healthy participants, we demonstrated that in Taiwanese participants the processes of thinking about oneself are more strongly associated with the anterior (prefrontal) hub of the DMN, whereas in Russian participants these processes were more strongly associated with the posterior (parietal) hub of this network. We hypothesized that representatives of Asian culture are more likely to have collectivist ethical attitudes, while representatives of Russian culture are more prone to individualism, which may be one of the reasons for the differences in resting brain activity. In our new study, we compare power spectral densities of EEG recorded between episodes of watching a video of a participant's own face, watching a video of a stranger's face, and watching a blank computer screen. Healthy Russian students and students from the People's Republic of China studying in Russia were invited as comparison groups. Indicators of collectivism, assessed using psychological questionnaires, as well as the age and gender of the subjects, were included in the analysis of EEG data as additional variables.

Materials and Methods:

Participants: 30 healthy Russian undergraduate and graduate students (mean age 23.4 years, 13 men, 17 women) and 30 Chinese undergraduate and graduate students studying at Novosibirsk State University (mean age 22.3 years, 16 men, 14 women) participated in the EEG experiment. All students filled out a set of psychological questionnaires on collectivism-individualism, affiliation with family values, the degree of fixation of attention on oneself, the level of personal, situational and social anxiety, and a susceptibility to depression. During the experiment, participants sat in a light- and sound-isolated room. The EEG was recorded in three stages: (1) 12 minutes of recording with repeated two-minute intervals with eyes closed and open (a total of 6 minutes with eyes closed and 6 minutes with open eyes), looking at a black computer screen while eyes are opened; (2) 12 minutes of recording with repeated two-minute intervals with eyes closed and open, looking at a recording of one's own face while eyes are opened; (3) 12 minutes of recording with repeated two-minute intervals with eyes closed and open, looking at a recording of a face, belonging to a stranger of a same gender as the participant, while eyes are opened.

Data acquisition and analysis: Brain electrical activity was recorded through 128 EEG-channels + EOG + ECG montage. Signal was acquired using NVX-136 amplifier (Moscow, Russia) with 0.3–100 Hz band-pass, digitized at 1000 Hz, using Cz as a reference and AFz as a ground. Eye-moving artifacts were removed using ICA. Further analysis used EEG recording of 100–120 s with eyes closed between looking at an empty screen or one's own or stranger's face. The power spectral densitiy in different frequency ranges, its dependence on the experimental condition, the group of participants and their psychological characteristics was analyzed using the sLORETA software package.

Results: (1) Psychological comparisons revealed that Chinese participants were indeed characterized by an increased level of collectivism indicators in comparison with Russian participants in most of the questionnaires we used. However, an assessment of the interaction between gender and nationality factors showed that higher rates of collectivism are typical for Chinese males than for Russian men, while Chinese and Russian women differed significantly less on this indicator. Chinese and Russian subjects did not differ in the degree of susceptibility to depression and the level of personal anxiety. (2) A comparison of the three experimental conditions in both groups of participants revealed that viewing one's own face compared to viewing a blank screen caused a decrease in power spectral density in the frontal cortex in the delta, alpha and gamma frequency ranges, while viewing another's face compared to observing a blank screen caused a decrease in power spectral density in the frontal, parietal and lateral temporal cortex in the same ranges as observing one's own face. In general, differences in power spectral density under different experimental conditions were observed in the cortical structures that are part of the default system. (3) Between-group differences between Russian and Chinese participants were most pronounced in the resting state after observing someone else's face and, to a lesser extent, after observing one's own facial video. Chinese participants showed greater power spectral density in alpha-band in the frontal cortex. (4) Collectivism scores were associated with greater alpha-band power spectral density in the frontal cortex in both groups of participants. This effect was significant after observing one's own face in Russian participants and after observing both one's own and another's face among Chinese participants. (5) All women, in comparison with men, regardless of their nationality, had an increased power spectral density in alpha-band in the parietal and lateral temporal cortex after observing their own face. Such gender differences were absent during resting state after viewing a blank screen or someone else's face.

Conclusions: We have proposed an experimental model that allows us to compare the electrical activity of the brain during resting state, induced by previous concentration of attention on stimuli related to the subject himself, another person, or an inanimate object. This model was used to study the dependence of resting brain activity on the sociocultural background of survey participants. Analysis of spectral density during resting state following observation of one's own facial video, facial video of a stranger, or a blank screen revealed features of a person's functional states, depending on his gender, ethnocultural background, and level of collectivism.

Funding: The study was supported by the Russian Science Foundation grant No. 22-15-00142 "fMRI and EEG correlates of self-focus as a predisposition factor to affective disorders". Work by A.E. Saprygin and A.N. Savostyanov on the development of methods for analyzing neurophysiological data was financed from the budget project of ICG SB RAS No. FWNR-2022-0020.

Список литературы/References

1. Bressler S.L., Menon V. Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn Sci.* 2010;14(6):277-290. doi 10.1016/j.tics.2010.04.004
2. Northoff G., Heinzel A., De Greck M., Bermpohl F., Dobrowolny H., Panksepp J. Self-referential processing in our brain – a meta-analysis of imaging studies on the self. *Neuroimage*. 2006;31(1):440-457. doi 10.1016/j.neuroimage.2005.12.002
3. Menon V. Large-scale brain networks and psychopathology: a unifying triple network model. *Trends Cogn Sci.* 2011;15(10):483-506. doi 10.1016/j.tics.2011.08.003
4. Knyazev G.G., Savostyanov A.N., Volf N.V., Liou M., Bocharov A.V. EEG correlates of spontaneous self-referential thoughts: a cross-cultural study. *Int J Psychophysiol*. 2012;86(2):173-181. doi 10.1016/j.ijpsycho.2012.09.002

doi 10.18699/bgrs2024-10-10

ЭЭГ-корреляты влияния спелеоклимата на мозговую активность взрослого здорового человека

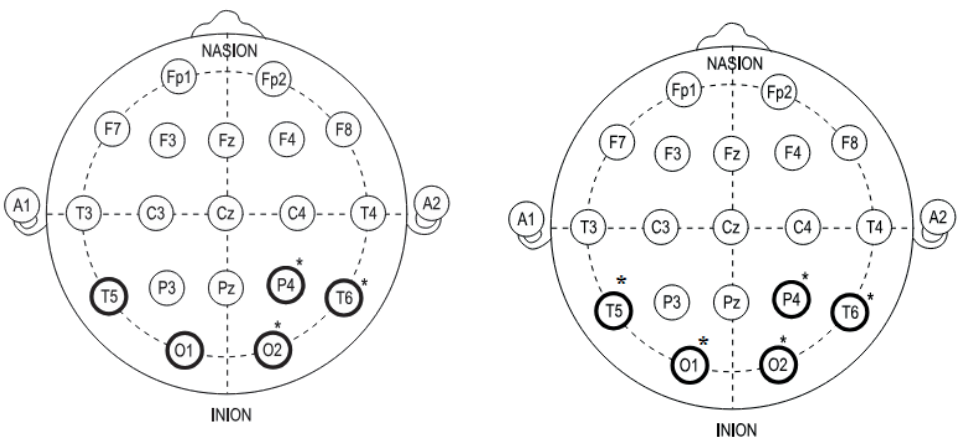
Семилетова В.А.
Воронежский государственный медицинский университет, Воронеж, Россия
vera2307@mail.ru

Ключевые слова: электроэнцефалограмма; ритм; индекс; спелеоклимат; спелеотерапия; аэроион

Мотивация и цель: Электроэнцефалография (ЭЭГ) – один их наиболее часто используемых и чувствительных методов оценки функциональной активности головного мозга. На ЭЭГ отражаются процессы утомления, стрессорного воздействия, изменения физических параметров среды – света, звука, ионного состава воздуха [1–5] и пр. Цель нашей работы: поиск ЭЭГ-коррелятов влияния спелеоклимата на мозговую активность взрослого здорового человека.
Методы и алгоритмы: В исследовании приняли участие 25 студентов-добровольцев 2 курса ВГМУ им. Н.Н. Бурденко, правши, в возрасте18–20 лет (15 человек экспериментальной группы и 10 человек контрольной группы; рассчитанная минимальная выборка для нашего исследования – 10 человек (по формуле Лера)). Критериями включения в экспериментальную группу явились: отсутствие острого периода вирусной или бактериальной инфекции, отсутствие обострения хронических заболеваний, а также анатомических изменений носовых ходов. Исследование соответствовало этическим стандартам. Каждый участник был проинформирован о цели исследования, подписал информированное согласие на участие в эксперименте. У студентов экспериментальной группы регистрация ЭЭГ проведена до и после 10-дневного часового курса спелеотерапии, в состоянии функционального покоя, наземной сильвинитовой спелеокамере 14 м^3, с помощью электроэнцефалографа-регистратора “Энцефалан – ЭЭГР – 19/26” фирмы Медиком (г. Таганрог) в 19 стандартных отведениях согласно международной схеме 10−20. Согласно замерам при выполнении исследований, температура воздуха в спелеокамере составляла 18–20 °С, относительная влажность воздуха – 65 %, содержание отрицательных аэроионов – 987 е/см^3, положительных аэроионов – 834 е/см^3, радиационныйфон – 17 мкР/час. У студентов контрольной группы регистрация ЭЭГ проведена до и после 10-дневного ежедневного часового отдыха, также в состоянии функционального покоя, с помощью электроэнцефалографа-регистратора «Энцефалан – ЭЭГР – 19/26» фирмы Медиком (г. Таганрог) в 19 стандартных отведениях согласно международной схеме 10−20. Проанализирована динамика следующих соотношений ритмов ЭЭГ (индексов): [альфа/бета1], [альфа/тета], [(тета + альфа)/бета1], хорошо зарекомендовавших себя в оценке активной умственной работы человека [6, 7]. Анализ полученных данных проведен с помощью программ Excel и IBM SPSS Statistics 26. Определена нормальность распределения признаков с использованием критерия Шапиро–Уилка, расчет достоверности отличий между зависимыми переменными проведен с использованием непараметрического критерия Уилкоксона для зависимых переменных.

Результаты: Статистический анализ изменений индексов [альфа/тета] не выявил значимых отличий до и после отдыха в контрольной группе. При этом слева отмечена тенденция к снижению медиан альфа/тета индексов в большинстве проанализированных отведений, справа подобного однонаправленного изменения не выявлено. В контрольной группе также выявлены статистически значимые изменения индексов [альфа/бета1] и [(тета + альфа)/бета1] в затылочной области справа под влиянием отдыха. Отмечено, что под влиянием ежедневного часового отдыха наблюдалось усиление медленно-волновой активности, коррелятами изменения мозговой активности под влиянием отдыха можно назвать индексы [альфа/бета1] и [(тета + альфа)/бета1].

В экспериментальной группе значимых отличий соотношений [(тета + альфа)/бета1] и [альфа/бета1] до и после курса спелеотерапии не выявлено, хотя наблюдалось увеличение данных индексов после 10-дневного воздействия спелеоклимата по отношению к исходному уровню. Отмечены значимые увеличения ($p<0.05$) индекса [альфа / тета] под воздействием спелеоклимата в затылочных отведениях справа (до: Q1=1.92; Me=2.80; Q3=4.83; после: Q1=3.65; Me=6.27; Q3=10.73) и слева (до: Q1=2.64; Me=6.25; Q3=14.39; после: Q1=1.43; Me=3.39; Q3=6.61), височных (T5 и T6) отведениях справа (до: Q1=5.46; Me=7.06; Q3=8.37; после: Q1=1.90; Me=3.17; Q3=5,03) и слева (до: Q1=1.17; Me=1.70; Q3=4.50; после: Q1=1.51; Me=6.04; Q3=11.84), теменном отведении P4 (до: Q1=1.28; Me=2.97; Q3=4.73 после: Q1=3.53; Me=5.36; Q3=7.40). Следует отметить, что в нашей более ранней работе по исследованию ЭЭГ человека при 2-часом воздействии спелеоклимата, выявлены значимые отличия также по индексу [альфа/тета] при отсутствии значимых отличий по другим исследованным соотношениям ритмов ЭЭГ, в тех же отведениях, рис. 1 [8].



Кратковременное воздействие спелеоклимата (2 ч)

10-дневный курс спелеотерапии

Рис. 1. Области изменения индексов альфа/тета под влиянием спелеоклимата (экспериментальная группа)

Полученные данные позволяют сделать вывод, что индекс [альфа/тета] можно считать наиболее чувствительным для оценки влияния как однократного

воздействия спелеотерапии на мозг человека, так и 10-дневного курса спелеотерапии.

EEG correlates of the influence of speleoclimate on the brain activity of an adult healthy person

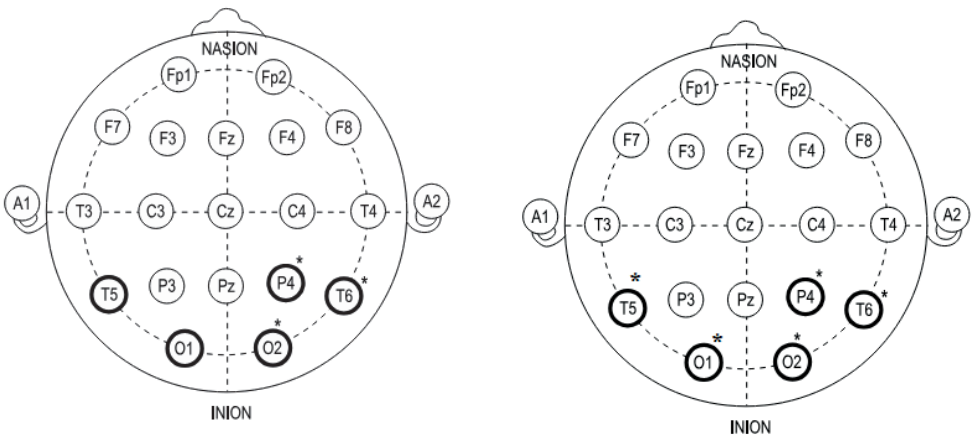
Semiletova V.A.
Voronezh State Medical University, Voronezh , Russia
vera2307@mail.ru

Key words: electroencephalogram; rhythm; index; speleoclimate; speleotherapy; aeroion

Motivation and Goal: Electroencephalography (EEG) is one of the most commonly used and sensitive methods for assessing the functional activity of the brain. The EEG reflects the processes of fatigue, stress, changes in the physical parameters of the environment – light, sound, ionic composition of air [1–5], etc. The purpose of our work: to search for EEG correlates of the influence of speleoclimate on the brain activity of an adult healthy person.
Methods and Algorithms: 25 volunteer 2nd year students of VSMU took part in the study, right-handed, 18–20 years old (15 people in the experimental group and 10 people in the control group; the calculated minimum sample for our study is 10 people (according to Lehr's formula)). The criteria for inclusion in the experimental group were: the absence of an acute period of viral or bacterial infection, the absence of exacerbation of chronic diseases, as well as anatomical changes in the nasal passages. The study complied with ethical standards. Each participant was informed about the purpose of the study and signed an informed consent to participate in the experiment. For students in the experimental group, EEG recording was carried out before and after a 10-day hour-long course of speleotherapy, in a state of functional rest, in a ground-based sylvinite speleochamber of 14 m^3 , using an electroencephalograph-recorder "Encephalan - EEGR – 19/26" from Medicom (Taganrog) in 19 standard leads according to the international 10−20 scheme. According to According to measurements during research, the air temperature in the caving chamber was 18–20 °C, relative air humidity was 65%, content of negative air ions – 987 e/cm^3, positive air ions – 834 e/cm^3 , background radiation – 17 μR/hour. For students in the control group, EEG was recorded before and after a 10-day daily one-hour rest, also in a state of functional rest, using an electroencephalograph-recorder "Encephalan – EEGR – 19/26" from Medicom (Taganrog) in 19 standard leads according to the international 10−20 scheme. The dynamics of the following EEG rhythm ratios (indices) were analyzed: [alpha/beta1], [alpha/theta], [(theta + alpha)/beta1], which have proven themselves in assessing active mental work of a person [6, 7]. The analysis of the obtained data was carried out using Excel and IBM SPSS Statistics 26. The normality of the distribution of characteristics was determined using the Shapiro-Wilk test, and the reliability of differences between dependent variables was calculated using the non-parametric Wilcoxon test for dependent variables.
Results: Statistical analysis of changes in indices [alpha/theta] did not reveal any significant differences before and after rest in the control group. At the same time, on the left there was a tendency towards a decrease in the median alpha/theta indices in the majority of the analyzed leads; on the right, no such unidirectional change was detected. In the control group, statistically significant changes in the indices [alpha/beta1] and

[(theta + alpha)/beta1] in the occipital areas on the right under the influence of rest. It was noted that under the influence of daily hour-long rest, an increase in slow-wave activity was observed; the indices [alpha/beta1] and [(theta + alpha)/beta1] can be called correlates of changes in brain activity under the influence of rest.

In the experimental group there were significant differences in the ratios [(theta + alpha)/beta1] and [alpha/beta1] before and after the course of speleotherapy was not detected, although an increase in these indices was observed after 10 days of exposure to the speleoclimate in relation to the initial level. There were significant increases ($p<0.05$) in the [alpha/theta] under the influence of speleoclimate in the occipital leads on the right (before: Q1=1.92; Me=2.80; Q3=4.83; after: Q1=3.65; Me=6.27; Q3=10.73) and on the left (before: Q1=2.64; Me=6.25 ; Q3=14.39; after: Q1=1.43; Q3=6.61), temporal (T5 and T6) leads (before: Q1=5.46; Q3=8.37; after: Q1=1.90; Me =3.17; Q3=5.03) and left (before: Q1=1.17; Me=1.70; Q3=4.50; after: Q1=1.51; Me=6.04; Q3=11.84), parietal lead P4 (before: Q1=1.28 ; Me=2.97; Q3=4.73 after: Q1=3.53; Q3=7.40). It should be noted that in our earlier work on the study of human EEG during 2-hour exposure to speleoclimate, significant differences were also revealed in the index [alpha/theta] in the absence of significant differences in other studied relationships of EEG rhythms, in the same leads, Fig. 1 [8].



Short-term exposure to speleoclimate (2 h)

10-day speleotherapy course

Fig. 1. Areas of change in alpha/theta indices under the influence of speleo climate (experimental group)

The data obtained allow us to conclude that index [alpha / theta] can be considered the most sensitive for assessing the effect as a single exposure speleotherapy on the human brain, and a 10-day course of speleotherapy .

Список литературы/References

1. Поликанова И.С., Сергеев А.В. Влияние длительной когнитивной нагрузки на параметры ЭЭГ. *Национальный психол. журнал.* 2014;1:84-92. doi 10.11621/npj.2014.010910
   [Polikanova I.S., Sergeev A.V. The influence of long-term cognitive load on EEG parameters. *National Psychol. Magazine*. 2014;1:84-92. doi 10.11621/npj.2014.010910 (in Russian)]
2. CMI Brain Research. Center for medicalinformation. Направление электрофизиологии НЦИЛС. [дата обращения: 05.05.2024]. Доступ по ссылке: https:/cmi.to

[CMI Brain Research. Center for medical information. Direction of electrophysiology NCILS. [access date: 05/05/2024]. Access via link: https:/cmi.to]
3. Михайлова Н.Л. Особенности электрической активности правого и левого полушарий головного мозга у лиц с разным профилем моторной асимметрии и ее связь с состоянием сердца в покое и после физической нагрузки. *Ульяновский мед.-биол. журнал.* 2011;2:147-155
[Mikhailova N.L. Features of electrical activity of the right and left hemispheres of the brain in individuals with different profiles of motor asymmetry and its connection with the state of the heart at rest and after physical activity. *Ulyanovsk Medical-Biol. Magazine*. 2011;2:147-155 (in Russian)]
4. Лебедева Н.Н., Каримова Е.Д. Нейрофизиологические проявления состояния монотонии у операторов с различной межполушарной асимметрией альфа-активности. *Журнал высшей нервной деятельности.* 2014;64(4):428
[Lebedeva N.N., Karimova E.D. Neurophysiological manifestations of the state of monotony in operators with different interhemispheric asymmetries of alpha activity. *J Higher Nervous Activity*. 2014;64(4):428 (in Russian)]
5. Игошина Т.В., Счастливцева Д.В., Котровская Т.И., Бубеев Ю.А. Динамика ЭЭГ-паттернов при коррекции стресс реакций методом ингаляции ксенона. *Вестник восстан. медицины.* 2017;1:116-121
[Igoshina T.V., Schastlivtseva D.V., Kotrovskaya T.I., Bubeev Yu.A. Dynamics of EEG patterns during correction of stress reactions using xenon inhalation. *Messenger is Restored. Medicine.* 2017;1:116-121 (in Russian)]
6. Поликанова И.С., Коршунов А.В., Леонов С.В., Веракса А.Н. Ассоциация рецептора к дофамину второго типа (DRD2) с развитием утомления в результате длительной когнитивной нагрузки. *Национальный психол. журнал.* 2016;3:115-126. doi 10.11621/npj.2016.0314
[Polikanova I.S., Korshunov A.V., Leonov S.V., Veraksa A.N. Association of dopamine receptor type 2 (DRD2) with the development of fatigue as a result of prolonged cognitive load. *National Psychol. Magazine*. 2016;3:115-126. doi 10.11621/npj.2016.0314 (in Russian)]
7. Луценко Е.Л. Физиологическая психофизиология. Особенности межполушарной асимметрии индекса альфа-ритма у студентов. *Вестник психофизиологии.* 2013;2:34-40
[Lutsenko E.L. Physiological psychophysiology. Features of interhemispheric asymmetry of the alpha rhythm index of students. *Bulletin Psychophysiology*. 2013;2:34-40 (in Russian)]
8. Семилетова В.А., Дорохов Е.В., Булгакова Я.В. Влияние кратковременного воздействия спелеотерапии на соотношение ритмов электроэнцефалограммы здорового человека. *Экология человека.* 2023;3:223-230. doi 10.17816/humeco192536
[Semiletova V.A., Dorokhov E.V., Bulgakova Y.V. The influence of short-term exposure to speleotherapy on the ratio of electroencephalogram rhythms in a healthy person. *Human Ecology*. 2023;3:223-230. doi 10.17816/humeco192536 (in Russian)]

doi 10.18699/bgrs2024-10-11

Использование инструмента MaxQuant для поиска биомаркеров БАР

Смирнова Л.П.*, Серегин А.А., Дмитриева Е.М.

Научно-исследовательский институт психического здоровья, Томский научно-исследовательский медицинский центр РАН, Томск, Россия

** lpsmirnova@yandex.ru*

Ключевые слова: биполярное аффективное расстройство; протеом; масс-спектрометрия; биомаркеры; сыворотка крови

Мотивация и цель: В последнее время активно развиваются новые подходы к диагностике психических расстройств и поиску их биологических маркеров, в том числе с помощью протеомного анализа. Особую актуальность приобретает поиск периферических маркеров, которые могут быть использованы для диагностики и прогноза эффективности терапии (тераностики). Публикации по протеомным исследованиям больных с психическими расстройствами малочисленны и в основном представлены работами по шизофрении. Статьи по анализу протеома больных биполярным аффективным расстройством (БАР) малочисленны и представлены в основном работами на постмортальном материале. В ряде исследований сыворотки крови больных БАР выявлены белки, связанные с митохондриальной дисфункцией и с нарушением энергетического обмена, не являющиеся специфичными для БАР. Основной задачей нашего исследования были поиск биомаркеров, специфично характеризующих разные клинические разновидности БАР, и оценка возможности применения этих маркеров в клинической практике.

Методы и алгоритмы: В работе использовали сыворотку крови больных биполярно-аффективным расстройством (БАР), проходивших лечение в отделении аффективных состояний НИИ психического здоровья ТНИМЦ, со средним возрастом больных 39.7 ± 15.6 года, длительностью заболевания 12 [3.5; 15.2] лет. Группу контроля составили здоровые волонтеры, сопоставимые по полу и возрасту с основной группой. После очистки сыворотки от мажорных белков был использован метод разделения белков в ПААГ. Вырезанные белковые бенды были подвергнуты триписинолизу и экстрации полученных пептидов по стандартному протоколу. Необработанные масс-спектрометрические данные анализировали с помощью программного обеспечения MaxQuant. Поисковая система Andromeda использовалась для поиска спектров МС/МС в базе данных человека Uniprot (содержащей 90 482 записи), объединенной с 262 распространенными примесями и с обратными версиями всех последовательностей. Идентификация пептида проводилась на основе поиска с начальным отклонением массы иона-предшественника до 7 ppm. Допуск на массу фрагмента был установлен на уровне 20 ppm по шкале m/z. Для количественного анализа рассматривались только белки, количественно содержащие по меньшей мере два пептида.

Результаты: Программный пакет MaxQuant позволяет приблизительно оценить относительное содержание белков в исследуемом образце при помощи индекса интенсивности пиков – iBAQ и его нормализованного индекса LFQ (для исключения ошибок при сопоставлении нескольких образцов). При помощи данного

инструмента у больных БАР было идентифицировано более 1200 белков, большинство из которых принимают участие в регуляции транскрипции и дифференцировке клеток, а также являются регуляторами широкого спектра сигнальных путей. Самой представленной функцией идентифицированных белков оказалась регуляция иммунного ответа. Наибольшее значение LFQ показал белок Serotransferrin (iBAQ 1883900 LFQ 3330900000). Серотрансферрин может играть роль в регуляции клеточной подвижности и пролиферации. Показано, что у пациентов с БАР, ранее не принимавших антидепрессанты, наблюдается снижение уровня серотрансферина [1]. Представляет интерес в качестве антивоспалительного белка Alpha-1-microglobulin/bikunin precursor (ген AMBP iBAQ 233710, LFQ intensity 22179000). Он является антиоксидантом, работает во внутрисосудистом русле и может играть защитную роль при воспалении, предотвращая чрезмерную проницаемость эндотелия сосудов и ГЭБ. Но этот белок был уже выявлен при маниакальном эпизоде БАР [2]. Белок, показавший наименьший LFQ, – это Protein S100-A7 (ген S100A7, iBAQ 98804 LFQ 528540). Это нейроспецифичный белок, основной функцией которого является цитокин-опосредованная регуляция хемотаксиса гранулоцитов, моноцитов и Т-клеток при воспалении. Данный белок входит в группу уникальных для нервной ткани кальций-связывающих белков. Значительное повышение уровня этого белка в сыворотке уже было показано во время обострений биполярного аффективного расстройства (эпизоды мании и депрессии), но не во время фаз ремиссии заболевания [3].
Выводы: Таким образом, использование популярного поискового инструмента MaxQuant при работе с сывороткой крови выявляет в основном белки, широко представленные в крови. Для поиска специфичных белков, отражающих патогенетические особенности БАР, необходимо использовать более чувствительные алгоритмы и программы.
Финансирование: Исследование поддержано грантом РНФ № 23-75-00023.

Using the MaxQuant tool to search for biomarkers of bipolar disorder

Smirnova L.P.*, Seregin A.A., Dmitrieva E.M.
Mental Health Research Institute, Tomsk National Medical Research Center, RAS, Tomsk, Russia
** lpsmirnova@yandex.ru*

Key words: bipolar disorder; proteome; mass spectrometry; biomarkers; blood serum

Motivation and Aim: Recently, new approaches to diagnosing mental disorders and searching for their biological markers, including using proteomic analysis, have been actively developing. Currently, the search for peripheral markers that can be used for diagnosis and prognosis of the effectiveness of therapy (theranostics) is of particular relevance. There are few publications on proteomic studies of patients with mental disorders, and are mainly represented by works on schizophrenia. Articles on the analysis of the proteome of patients with bipolar affective disorder (BD) are few in number and are represented mainly by works on postmortem material. A number of studies of the blood serum of patients with BD have identified proteins associated with

mitochondrial dysfunction and impaired energy metabolism that are not specific to BD. The main objective of our study was to search for biomarkers that specifically characterize different clinical types of BD and to assess the possibility of using these markers in clinical practice.

Methods and Algorithms: The work used the blood serum of patients with bipolar disorder (BD) who were treated in the department of affective states of the Scientific Research Institute of Mental Health, TNRMC, with an average age of patients of 39.7 ± 15.6 years, duration of the disease 12 [3.5; 15.2] years. The control group consisted of healthy volunteers comparable in gender and age to the main group. After purifying the whey from major proteins, the PAGE separation method was used. The excised protein bands were subjected to trypisinolysis and extraction of the resulting peptides according to a standard protocol. The mass spectrometric raw data were analyzed with the MaxQuant software. The Andromeda search engine was used to search the MS/MS spectra against the Uniprot human database (containing 90,482 entries) combined with 262 common contaminants and concatenated with the reversed versions of all sequences. Peptide identification was based on a search with an initial mass deviation of the precursor ion of up to 7 ppm. The fragment mass tolerance was set to 20 ppm on the m/z scale. Only proteins quantified with at least two peptides were considered for quantitation.

Results: The MaxQuant software package allows you to approximately estimate the relative abundance of proteins in the test sample using the peak intensity index – iBAQ and its normalized LFQ index (to eliminate errors when comparing several samples). Using this tool, more than 1200 proteins were identified in BD patients, most of which are involved in transcription regulation and cell differentiation, and are also regulators of a wide range of signaling pathways. The most prominent function of the identified proteins was the regulation of the immune response. The highest LFQ value was shown by the protein Serotransferrin (iBAQ 1883900 LFQ 3330900000). Serotransferrin may play a role in the regulation of cell motility and proliferation. It has been shown that in patients with BD who have not previously taken antidepressants, there is a decrease in serotransferrin levels [1]. Of interest as an anti-inflammatory protein is Alpha-1-microglobulin/bikunin precursor (gene – AMBP iBAQ 233710, LFQ intensity 22179000). It is an antioxidant, works intravascularly and may play a protective role in inflammation by preventing excessive permeability of the vascular endothelium and the BBB. But this protein has already been identified in manic episodes of BD [2]. The protein that showed the lowest LFQ is the Protein S100-A7 protein (gene – S100A7, iBAQ 98804 LFQ 528540). This is a neurospecific protein, the main function of which is cytokine-mediated regulation of the chemotaxis of granulocytes, monocytes, and T cells during inflammation. This protein is part of a group of calcium-binding proteins unique to nervous tissue. A significant increase in serum levels of this protein has already been shown during exacerbations of bipolar disorder (episodes of mania and depression), but not during remission phases of the disease [3].

Conclusion: Thus, the use of the popular MaxQuant search tool when working with blood serum mainly identifies proteins that are widely present in the blood. To search for specific proteins that reflect the pathogenetic features of bipolar disorder, it is necessary to use more sensitive algorithms and programs.

Funding: The study is supported the Russian Science Foundation No. 23-75-00023.

Список литературы/References

1. Stelzhammer V., Haenisch F., Chan M.K., Cooper J.D., Steiner J., Steeb H., Martins-de-Souza D., Rahmoune H., Guest P.C., Bahn S. Proteomic changes in serum of first onset, antidepressant drug-naïve major depression patients. *Int J Neuropsychopharmacol*. 2014;17(10):1599-608. doi 10.1017/S1461145714000819
2. Cerit C., Sarıhan M., Nart Ö., Kasap M., Yaşar H., Akpınar G. Are Mannanbinding Lectine Serin Protease-2 and Alpha-1-microglobulin and Bukinin Precursor the Potential Biomarkers of Manic Episode? A Study via Urinary Proetomic Analysis. *Clin Psychopharmacol Neurosci*. 2021;19(2):269-281. doi 10.9758/cpn.2021.19.2.269
3. Wolf R., Howard O.M., Dong H.F. et al. Chemotactic activity of S100A7 (Psoriasin) is mediated by the receptor for advanced glycation end products and potentiates inflammation with highly homologous but functionally distinct S100A15. *J Immunol*. 2008;181(2):1499-506. doi 10.4049/jimmunol.181.2.1499

doi 10.18699/bgrs2024-10-12

Апробация протокола и электрофизиологические изменения, ассоциированные с многозадачным когнитивным тренингом в виртуальной реальности

Тарасова И.*, Куприянова Д., Кухарева И., Соснина А., Ляпина И., Трубникова О., Барбараш О.

Научно-исследовательский институт комплексных проблем сердечно-сосудистых заболеваний, Кемерово, Россия

** iriz78@mail.ru*

Ключевые слова: когнитивный тренинг; ЭЭГ; альфа-активность; виртуальная реальность

Обоснование и цель: Нейротехнологии, такие как виртуальная реальность (ВР), имеют большой потенциал для медицинских исследований и практики, позволяя расширить возможности реабилитационных мероприятий [1, 2]. Технология ВР создает мультисенсорную и мультидоменную стимуляцию когнитивных функций [3, 4]. В связи с этим использование ВР может быть эффективным методом когнитивной реабилитации для пациентов с различной мозговой патологией, в том числе острой или хронической ишемией мозга [5]. Разработка и апробация протоколов когнитивного тренинга (КТ) с использованием ВР, как и изучение нейрофизиологических изменений, ассоциированных с ним, проводится в основном на практически здоровых лицах, с дальнейшим расширением результатов исследований на целевую группу пациентов. Целью данной работы явились апробация разработанного оригинального протокола и исследование нейрофизиологических эффектов ВР при проведении многозадачного КТ у здоровых испытуемых.

Материалы и методы: Для изучения эффективности многозадачного КТ с использованием ВР был разработан специализированный программно-аппаратный комплекс. С использованием оригинального тренировочного протокола комплекс ВР был впервые протестирован на группе из 25 практически здоровых лиц (из них 15 женщин); средний возраст 28.1±10.6 года). Участники оценивали комплекс ВР с помощью опросников, в том числе System usability scale (SUS) [6], сразу же после проведения эксперимента для определения интереса и удобства его использования. Стандартное неврологическое обследование, психометрическое тестирование, а также ЭЭГ исследования были проведены до и после тестирования оригинального протокола многозадачного КТ с использованием ВР.

Результаты: Участники продемонстрировали высокую степень принятия предполагаемого будущего использования, отношения и удовольствие от тренинга. Среднее значение шкалы SUS 72.6±11.9 балла (макс. 100). У части испытуемых (6/25) наблюдались незначительные эффекты векции после проведения многозадачного КТ с использованием ВР. Также после проведения эксперимента наблюдалось статистически значимое улучшение внимания и пространственных навыков ($p < 0.05$). По данным ЭЭГ было обнаружено статистически значимое увеличение мощности альфа-ритма по сравнению с показателями перед проведением эксперимента ($p < 0.0001$). Кроме того, субъекты продемонстрировали незначительное снижение мощности биопотенциалов тета-ритма.

Соотношение тета/альфа значительно уменьшилось после тестирования многозадачного КТ с использованием ВР по сравнению с исходными значениями ($p < 0.0001$). Было обнаружено, что постэкспериментальные логарифмированные показатели мощности альфа-ритма имеют положительную корреляцию с индикаторами задачи умственного вращения и исполнительных функций после тестирования многозадачного КТ с использованием ВР.
Выводы: Полученные данные свидетельствуют о том, что разработанный комплекс для многозадачного КТ с использованием ВР и протокол тренинга являются приемлемыми и удобными для использования практически здоровыми людьми. Нейрофизиологические эффекты многозадачного КТ с использованием ВР у здоровых испытуемых продемонстрировали ингибирование сенсорных процессов и подавление отвлекающих факторов как проявление общего тормозящего механизма в мозге при когнитивной нагрузке. Количественные исследования ЭЭГ необходимы для подтверждения доступности многозадачного КТ с использованием ВР для пациентов с когнитивными расстройствами, в том числе послеоперационной когнитивной дисфункцией.
Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 23-15-00379 (https://rscf.ru/project/23-15-00379/).

Protocol testing and electrophysiological changes associated with multitasking cognitive training in virtual reality

Tarasova I.*, Kupriyanova D., Kukhareva I., Sosnina A., Lyapina I., Trubnikova O., Barbarash O.

Research Institute for Complex Issues of Cardiovascular Diseases, Kemerovo, Russia
** iriz78@mail.ru*

Key words: cognitive training; EEG; alpha activity; virtual reality

Background and Aim: Neurotechnologies such as virtual reality (VR) have a great potential for medical research and practice, making it possible to expand rehabilitation activities [1, 2]. VR technology creates multisensory and multidomain stimulation of cognitive functions [3, 4]. Therefore, VR can be an effective method of cognitive rehabilitation for patients with various brain pathologies, including acute or chronic brain ischemia. [5] The development and testing of cognitive training (CT) protocols using VR, as well as the study of neurophysiological changes associated with it, is carried out mainly on practically healthy individuals, with the further extension of the research results to the target group of patients. The aim of this work was to evaluate the original protocol and examine the physiological effects of VR during multitasking CT in healthy subjects.
Materials and Methods: A specialized hardware and software complex was created to investigate the effectiveness of multitasking CT using VR. Using the original training protocol, the VR complex was tested for the first time on a group of 25 practically healthy individuals (including 15 women); the average age is 28.1±10.6 years). The VR complex was evaluated by participants with questionnaires, including the System usability scale (SUS) [6], immediately after the experiment to determine interest and usability. Standard neurological examinations, psychometric testing, and EEG studies were conducted before and after testing the original multitasking CT protocol using VR.

Results: All participants were highly accepting of the intended future use, attitude, and pleasure of training. Average SUS score was 72.6±11.9 (max. 100). Some of the subjects (6/25) experienced minor effects of vection after conducting a multitasking CT using VR. There was also a statistically significant improvement in attention and spatial skills after multitasking CT using VR ($p<0.05$). EEG data revealed a significant increase in alpha power compared to pre-experiment levels ($p<0.0001$). Additionally, the subjects indicated a slight decrease in theta power. The theta/alpha ratio has significantly decreased after the VR multitasking CT as compared to the baseline ($p<0.0001$). Post-experimental log alpha power scores have been found to be positively correlated with indicators of the mental rotation task and executive functions after VR multitasking CT. *Conclusion*: The obtained data shows that the developed complex for multitasking CT using VR with the original training protocol is acceptable and convenient for use by practically healthy people. Neurophysiological effects of multitasking CT using VR in healthy subjects have demonstrated sensory inhibition and distraction suppression as a manifestation of the overall inhibitory mechanism in the brain under cognitive load. Quantitative EEG studies are needed to confirm the availability of VR-based multitasking CT for patients with cognitive disorders, including postoperative cognitive dysfunction.

Funding: This research was funded by the Russian Science Foundation No. 23-15-00379, https://rscf.ru/en/project/23-15-00379/.

Список литературы/References

1. Venkatesan M., Mohan H., Ryan J.R. et al. Virtual and augmented reality for biomedical applications. *Cell Rep Med*. 2021;2:100348
2. Tieri G., Morone G., Paolucci S., Iosa M. Virtual reality in cognitive and motor rehabilitation: Facts, fiction and fallacies. *Expert Rev Med Devices*. 2018;15:107-117
3. Bauer A.C., Andringa G. The potential of immersive virtual reality for cognitive training in elderly. *Gerontology*. 2020;66:614-623
4. Wenk N., Penalver-Andres J., Buetler K.A., Nef T., Müri R.M., Marchal-Crespo L. Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment. *Virtual Real*. 2023;27:307-331
5. Saeedi S., Ghazisaeedi M., Rezayi S. Applying game-based approaches for physical rehabilitation of post-stroke patients: A systematic review. *J Healthc Eng*. 2021;2021:9928509
6. Cheah W.H., Mat Jusoh N., Aung M.M.T., Ab Ghani A., Mohd Amin Rebuan H. Mobile Technology in Medicine: Development and Validation of an Adapted System Usability Scale (SUS) Questionnaire and Modified Technology Acceptance Model (TAM) to Evaluate User Experience and Acceptability of a Mobile Application in MRI Safety Screening. *Indian J Radiol Imaging*. 2022;33(1):36-45. doi 10.1055/s-0042-1758198

doi 10.18699/bgrs2024-10-13

Замечания к проекту фундаментальной теории разума

Филимонов В.*

Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия
** filimonov-v-a@yandex.ru*

Ключевые слова: методология; логическая «клеточка»; разум; интеллект; определения

Мотивация и цель: В 2015 г. в докладе [1] была представлена впечатляющая программа создания фундаментальных теорий разума и мозга. Появились публикации с описанием значительных результатов, такие как [2], однако до настоящего времени у автора отсутствуют сведения о продвижении в реализации самой программы. Причиной, скорее всего, являются методологические аспекты. Целью работы является обсуждение некоторых из них.
Методы и алгоритмы: Использован простой системный анализ (Ф.П. Тарасенко, Р. Акофф), а также методы А.А. Зиновьева [3], В.А. Лефевра [4–6] и авторская схема «4Ф» [7].
Результаты: Поставленная в [1] задача соответствует подходу Московского методологического кружка, в котором использовалась структура исследования вида «*Процесс/Материал/Реализация Процесса на Материале*». В данном случае это структура «*Разум/Носитель/Реализация Разума на Носителе*». Для построения теории целесообразно сформировать логическую (по аналогии с биологической) «клеточку» – наиболее простую структуру, содержащую основные свойства целого [3]. В связи со сложностью исследуемого объекта (процесса) такая «клеточка» должна содержать модель самого Исследователя [5], причем в виде фундаментальной триады «Исследователь/Логика (=конструктор Исследователя)/Онтология (=модель исследуемого объекта для данного конструктора)» [3]. Также представляется целесообразным для моделирования катастрофических фрагментов процесса иметь в «клеточке» не менее двух джокеров (точек случайного разветвления в синергетике): одну для объекта и одну для модели Исследователя.
Придерживаясь подхода В.А. Лефевра [4], автор не считает мозг единственным материалом в качестве носителя разума. Например, модельный организм *Physarum polycephalum*, не имеющий мозга и представляющий собой колонию клеток, демонстрирует подверженность эффекту асимметричного доминирования [8, с. 458–466]. Такое поведение является вполне рациональным (или разумным, что зависит от определения разума; см. ниже). Пленочные колонии *Bacillus subtilis* могут проявлять «чувство кворума» и координируют свои действия не только химическим путем, но и с помощью электрических сигналов [9]. Применительно к системам искусственного интеллекта (ИИ) напомним, что, помимо цифровых компьютеров, существует достаточно много аналоговых вариантов: электрические, гидравлические, пневматические, квантовые. Таким образом, носителю разума достаточно обеспечивать какой-либо вариант сетевого взаимодействия (коммуникации). С другой стороны, одушевленные существа являются альгедоническими системами, что предполагает реакцию всего организма, а не только мозга.

Для определения понятия «разум» можно использовать квази-остенсивные определения, т. е. указать примеры разумного поведения, а также результаты работы разума. Одним из таких примеров является «изобретение» системой ИИ в 2018 г. аналога Периодической системы химических элементов (Таблица Д.И. Менделеева, 1869 г.) [10]. Однако представляется маловероятным, что ИИ сможет без соответствующей подсказки получить аналог следующей версии – Горно-Алтайской таблицы химических мультиплетов KuFeRum (Ю.И. Кулаков, 1999 г.).

Относительно понятия «сеть» можно предложить как минимум два варианта определения. Первое определение относится к объектам, которые являются сетями онтологически, независимо от Исследователя. Таковы, например, рыболовная сеть и сеть нейронов мозга. Эти сети можно определить как структуры, образованные пересечениями линий. Второе определение относится к тем объектам, которые Исследователь представляет в своей Логике как сети. Таковы сети соавторов публикаций и социальные сети.

Автор предпочитает определения, сделанные по формату формулы изобретения, который применяется в патентном деле: «Предлагается *, предназначенный для **, состоящий из ***, ОТЛИЧАЮЩИЙСЯ тем, что с целью **** произведены следующие изменения: *****». Для лучшего понимания может быть использована приведенная в табл. 1 схема «4Ф» (Фамильное имя – Формы – Функции – Фундамент) [8]. Поясним, что термином “familia” в Древнем Риме назывались все рабы, принадлежащие конкретному владельцу.

Таблица 1. Два примера использования схемы «4Ф»

Уровень	Описание на этом уровне	
Фамильное имя (варианты, синонимы, близкие понятия)	**Часы** Клепсидра, брегет, ходики куранты, хронометр, …	**Разум** Творчество, креативность, сознание, ум, рассудок, интеллект, логика, гениальность, здравый смысл, интуиция, трансцендентное знание, …
Формы реализации и проявления (обнаружения)	Песочные, кварцевые, солнечные, атомные, … Наручные, круглые, … Мужские, женские, …	Поведение. Результат мыслительной работы. Сеть, гиперсеть, …
Функции в деятельности	Измерение времени. Украшение, Индикатор статуса, …	Обеспечение эволюции. Развитие сфер (культура, искусство, наука, религия, …)
Фундамент Законы, позволяющие объекту /процессу выполнять функции в данной форме	Астрономические и физические закономерности. Цикличность. Понятия «раньше/позже», Социальные законы	Формирование (гипер)сети и процесс движения по этой сети. Язык

Идея фирмы IKEA в предоставлении возможности потребителю самому участвовать в сборке изделия увеличивает привлекательность такого изделия. Аналогично можно дать возможность разуму, как искусственному, так и

естественному, участвовать в создании модели себя. Такая работа потребует от Исследователя определенной квалификации в сфере психотерапии.
Выводы: Для продвижения в построении фундаментальной теории разума необходимо сформировать логическую «клеточку» этого понятия без привязки к его конкретному носителю (в частности, мозгу). В состав такой клеточки должен входить сам Исследователь со своим исследовательским арсеналом, а также компоненты, отражающие возможные катастрофические ситуации. В рамках этой теории могут быть сконструированы несколько определений понятия «разум» в зависимости от способов их использования.
Финансирование: Работа выполнена в рамках государственного задания ИМ СО РАН, проект FWNF-2022-0016.

Remarks on the draft of the fundamental theory of mind

Filimonov V.*
Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia
** filimonov-v-a@yandex.ru*

Key words: methodology; logical “cell”; mind; intelligence; definitions

Motivation and Aim: In 2015, the report [1] presented an impressive program for creating fundamental theories of the mind and brain. There have been publications describing significant results, such as [2], but so far the author has no information about progress in the implementation of the program itself. The reason is most likely methodological aspects. The purpose of the paper is to discuss some of them.
Methods and Algorithms: A simple system analysis was used (F.P. Tarasenko, R. Ackoff), as well as the methods of A.A. Zinoviev [3], V.A. Lefebvre [4–6] and the author's scheme “4F” [7].
Results: The task set in [1] corresponds to the approach of the Moscow Methodological Circle, which used a research structure of the type “Process/Material/Implementation of the Process on the Material”. In this case, it is the structure of the “Mind/Carrier/The realization of the Mind on the Carrier”. To build a theory, it is advisable to form a logical (by analogy with a biological) “cell” – the simplest structure containing the basic properties of the whole [3]. Due to the complexity of the object (process) under study such a “cell” should contain a model of the Researcher himself [5], and in the form of a fundamental triad “Researcher / Logic (=Researcher's constructor) / Ontology (=model of the object under study for this constructor)” [3]. It also seems advisable to have at least two jokers (points of random branching in synergetic) in the “cell” for modeling catastrophic fragments of the process: one for the object and one for the Researcher's model.
Adhering to the approach of V.A. Lefebvre [4], the author does not consider the brain to be the only material as a carrier of the mind. For example, the model organism *Physarum polycephalum*, which does not have a brain and is a colony of cells, demonstrates susceptibility to the effect of asymmetric dominance [8, p. 458–466]. This behavior is quite rational (or reasonable, depending on the definition of mind; see below). *Bacillus subtilis* film colonies can exhibit a “sense of quorum” and coordinate their actions not only chemically, but also using electrical signals [9]. With regard to artificial intelligence (AI) systems, we recall that, in addition to digital computers, there are quite a lot of

analog options: electric, hydraulic, pneumatic, quantum. Thus, it is enough for the mind carrier to provide some kind of network interaction (communication) option. On the other hand, animate beings are algedonic systems, which implies the reaction of the whole organism, not just the brain.

To define the concept of "mind", quasi-intensive definitions can be used, i. e., examples of reasonable behavior can be indicated, as well as the results of the mind's work. One such example is the "invention" by the AI system in 2018 of an analogue of the Periodic Table of Chemical Elements (Table of D.I. Mendeleev, 1869) [10]. However, it seems unlikely that AI will be able to obtain an analogue of the next version without an appropriate hint – the Gorno-Altai table of chemical multiplets KuFeRum (Yu.I. Kulakov, 1999)ю

Regarding the concept of "network", at least two definitions can be proposed. The first definition refers to objects that are networks ontologically, regardless of the Researcher. This is, for example, a fishing net and a network of brain neurons. These networks can be defined as structures formed by intersections of lines. The second definition refers to those objects that the Researcher represents in his Logic as networks. These are the networks of co-authors of publications and social networks.

The author prefers definitions made according to the format of the claims, which is used in the patent case: "It is proposed *, intended for **, consisting of ***, HAVING DIFFERENCES in that the following changes were made for the purpose of ****: *****". For a better understanding, the "4F" scheme (Family Name – Forms – Functions – Foundation) shown below in Table 1 can be used [8]. Let us explain that the term "familia" in Ancient Rome referred to all slaves belonging to a specific owner.

Table 1. Two examples of using the "4F" scheme

Level	Description at this level	
Family name (variants, synonyms, similar concepts)	**Timepiece** Clepsydra, breguet, chimes, chronometer, clock, watch, …	**Mind** Creativity, consciousness, reason, intelligence, logic, genius, common sense, intuition, transcendental knowledge, …
Forms of realization and manifestation (detection)	Sand, Quartz, Solar, Atomic, … Wrist watch, round, … Men's, women's, …	Behaviour. The result of mental work. Network, hypernet, …
Functions in the activity	Time measurement. Decoration, Status indicator	Ensuring evolution. The development of spheres (culture, art, science, religion, ...)
Foundation Laws that allow an object or process to perform the functions in these forms	Astronomical and physical laws. Cyclicity. The concepts of "earlier/later". Social laws	Formation (hyper) networks and the process of movement on this networks. Language

IKEA's idea of enabling the consumer to participate in the assembly of the product himself increases the attractiveness of such a product. Similarly, it is possible to enable the mind, both artificial and natural, to participate in creating a model of oneself. Such work will require the Researcher to have certain qualifications in the field of psychotherapy.

Conclusion: To advance in the construction of a fundamental theory of mind, it is necessary to form a logical "cell" of this concept without reference to its specific carrier (in particular, the brain). Such a cell should include the Researcher himself with his research arsenal, as well as components reflecting possible catastrophic situations. Within the framework of this theory, several definitions of the concept of "mind" can be constructed, depending on the ways they are used.

Funding: The work was carried out within the framework of the state assignment of the SB RAS, project FWNF-2022-0016.

Список литературы/References

1. Анохин К.В. Когнитом: гиперсетевая структура мозга. https://vk.com/wall-74058720_1256?ysclid=lqw7gbk76o184584669&z=video-74058720_171048702%2F3c2f88dc09d5aece1d
   [Anokhin K.V. Cognitom: hypernetwork structure of the brain (video in Russian)]
2. Анохин К.В., Новоселов К.С., Смирнов С.К., Ефимов А.Р., Матвеев Ф.М. Искусственный интеллект для науки и наука для искусственного интеллекта. *Вопросы философии*. 2022;3:93-105. doi 10.21146/0042-8744-2022-3-93-105
   [Anokhin K.V., Novoselov K.S., Smirnov S.K., Efimov A.R., Matveev F.M. Artificial intelligence for science and science for artificial intelligence. *Problems of Philosophy*. 2022;3:93-105. doi 10.21146/0042-8744-2022-3-93-105 (in Russian)]
3. Зиновьев А.А. Логический интеллект. М.: Изд-во Московского гуманитарного ун-та, 2005
   [Zinoviev A.A. Logical intelligence. Moscow: Publishing House of the Moscow University of the Humanities, 2005 (in Russian)]
4. Лефевр В.А. Общая схема современной психологии. Место рефлексивных исследований в системе наук. *Математические структуры и моделирование*. 2018;1(45):108-110
   [Lefebvre V.A. The general scheme of modern psychology. The place of reflexive research in the system of sciences. *Mathematical Structures and Modeling*. 2018;1(45):108-110 (in Russian)]
5. Лефевр В.А. Конфликтующие структуры. М.: Советское радио, 1973
   [Lefebvre V.A. Conflicting structures. Moscow: Soviet Radio, 1973 (in Russian)]
6. Лефевр В.А. Что такое одушевленность? М: Когито-Центр, 2017
   [Lefebvre V.A. What is animacy? Moscow: Kogito-Center, 2017 (in Russian)]
7. Филимонов В.А., Чернявская В.С. Формализация одушевленности на примере понятия «любовь». *Онтология проектирования*. 2022;12(1):11-24. doi 10.18287/2223-9537-2022-12-1-11-24
   [Filimonov V.A., Chernyavskaya V.S. Formalization of animacy by the example of the concept of "love". *Ontology of Design*. 2022;12(1):11-24 (in Russian)]
8. Элленберг Дж. Как не ошибаться. Сила математического мышления. М.: МИФ, 2018
   [Ellenberg J. How not to be wrong. The power of mathematical thinking. Moscow: MIF, 2018 (in Russian)]
9. Ward L.D., Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res*. 2016;44(D1):D877-D881. doi 10.1093/nar/gkv1340
10. Zhou Q., Tang P., Liu S., Pan J., Yan Q., Zhang S.C. Learning atoms for materials discovery. *Proc Natl Acad Sci USA*. 2018;115(28):E6411-E6417. doi 10.1073/pnas.1801181115

doi 10.18699/bgrs2024-10-14

Когнитивные и личностные особенности сиблингов и звучащая речь

Шаляпина А.*, Абрамкина Е.
Новосибирский государственный университет, Новосибирск, Россия
** a.shalyapina@g.nsu.ru*

Ключевые слова: когнитивные механизмы; сиблинги; психологические личностные черты; звучащая речь; спонтанная речь; подготовленная речь; формантный анализ; PLS-анализ

Мотивация и цель: Влияние когнитивных механизмов на языковые и акустические характеристики речи диктора изучено не столь широко, как влияние возраста, пола, конфигурации речевого тракта. В случае сиблингов когнитивные механизмы могут оказаться схожими. Вопрос влияния родственных отношений на особенности речи сиблингов представляется интересным для рассмотрения еще и потому, что в жизни большинства людей сиблинговые отношения являются одними из самых длительных, следовательно, оказывающими огромное влияние на социализацию, личностное развитие. В данном пилотном исследовании были поставлены две цели: 1) описать когнитивные механизмы сиблингов с помощью понятийного аппарата психологических личностных черт; 2) оценить, какой из факторов, наиболее значимых и считываемых для сиблинговых групп (порядок рождения детей, пол, разница в возрасте), оказывает превалирующее влияние на особенности звучащей речи сиблингов.
Участники исследования: В пилотном исследовании приняли участие 8 человек (2 пары родных сестер, 2 пары родных братьев) в возрасте от 15 до 25 лет. Испытуемые заполнили опросник темперамента и характера, а также предоставили аудиозаписи спонтанной и подготовленной речи (общее количество аудиозаписей – 16 файлов формата .mp3). Длительность записей спонтанной речи (монолог на свободную тему) составила в среднем 3–4 минуты, подготовленной речи (чтение с листа художественного текста) – 5–7 минут (общая длительность проанализированных материалов составила около 80 минут).
Методы исследования: Опросник темперамента и характера Н. Алмаева и Д. Островской; формантный анализ (определение лингвистических, физических, акустических параметров речи дикторов); PLS-анализ (получение проекций на латентные структуры данных для анализа полученных данных и визуализации результатов).
Результаты: Для устной речи было установлено, что главным имплицитным механизмом является порядок рождения. Пол и возраст испытуемых либо влияют намного меньше, либо не влияют вообще (в контексте рассмотрения пар сиблингов).
Выводы: Речевое развитие напрямую связано с формированием личностных черт человека. Превалирующим фактором сиблингового влияния на форманты F1–F4 в парах одного пола является порядок рождения.

Cognitive and personality siblings' features and oral speech

Shaliapina A.*, Abramkina E.
Novosibirsk State University, Novosibirsk, Russia
** a.shalyapina@g.nsu.ru*

Key words: cognitive mechanisms; siblings; psychological personality traits; oral speech; spontaneous speech; prepared speech; formant analysis; PLS-analysis

Motivation and Aim: The influence of cognitive mechanisms on the linguistic and acoustic characteristics of speech has not been studied as much as the influence of age, gender, and speech tract configuration. In the case of siblings, cognitive mechanisms may be similar for speakers. The question of the influence of sibling relations on the speech features is interesting to consider also because of the nature of this type of relations: for most people sibling relations are one of the longest-lasting in their life, therefore, it has a great influence on personal development. The two aims of this pilot study were: (1) to describe cognitive mechanisms of siblings using the conceptual apparatus of psychological personality traits; (2) to assess which factors are most significant and readable for sibling groups (birth order, gender, age difference), which influence more on the siblings' speech features.
Participants: The pilot study involved 8 participants (2 pairs of brothers, 2 pairs of sisters) aged 15–25 years. They filled out a temperament and character questionnaire and recorded audio files of spontaneous and prepared speech (total number of audio recordings – 16 .mp3 files); the duration of the recordings of spontaneous speech (monologue, free topic) was 3–4 minutes (on average), of prepared speech (reading of fiction text) – 5–7 minutes (total duration of the analyzed materials is about 80 minutes).
Methods: temperament and character questionnaire by N. Almaev and D. Ostrovskaya; formant analysis (determination of linguistic, physical, acoustic parameters of speakers' speech); PLS-analysis (obtaining projections on latent data structures to analyze the obtained data and visualize the results).
Results: For oral speech, birth order is the main implicit mechanism, sex and age of participants either affecting much less or not at all (in considering of sibling pairs).
Conclusion: Speech development is directly related to the formation of human personality traits. The prevailing factor of sibling influence on F1–F4 formants in same-sex pairs is birth order.

doi 10.18699/bgrs2024-10-16

Associations of self-focus attention to oscillatory dynamics of self-recognition in morphed images

Bocharov A.[1, 2]*, Savostyanov A.[1, 2], Lebedkin D.[1, 2], Saprigyn A.[1], Tamozhnikov S.[1], Rudych P.[2], Merkulova E.[1], Knyazev G.[1]

[1] *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *bocharovav@neuronm.ru*

Key words: EEG; morphed faces; private self-consciousness; public self-consciousness; social anxiety; delta rhythm; theta rhythm

Motivation and Aim: Recognizing one's own face is a unique, complex process related to basic processes of self-awareness [1]. Self-focus has attracted the attention of researchers as an important trait predisposing to affective disorders [2, 3]. In this study, using the EEG recordings, we aimed to investigate the associations between the focus of attention on the self and the oscillatory dynamics accompanying the processes of recognizing morphed images consisted of self and stranger faces.

Methods: Forty-nine volunteers (30 females) between the ages of 18 and 42 participated in the study. Study participants completed the Self Consciousness Scale [4, 5], which contains the Private Self-Consciousness and Public Self-Consciousness/Social Anxiety subscales. During the 127-channel EEG recording, subjects were randomly presented morphed images of their own face with the face of a stranger. The face of a stranger of the same sex and similar attractiveness was selected from the NimStim Set of Facial Expressions database. It was needed to choose 1 of 2 options whether the person recognized himself or not in the morphed image by pressing the corresponding buttons. The stimulus presentation program was designed to select a morphing stage during testing at which the probability of recognizing oneself or the other was equal. Spearman correlation analyses were conducted at a significance level of $p < 0.005$ for the Private Self-Consciousness and the Public Self-Consciousness/Social Anxiety subscales with each point of the time-frequency resolution of event-related spectral perturbations separately for the conditions when the participant recognized him/herself and in the condition when he/she did not.

Results and Conclusion: In the condition of not recognizing oneself in the morphed image, the Private Self-Consciousness subscale was positively related to delta rhythm synchronization. According to the literature on the relation of the delta rhythm increase in arousal and motivation states [6, 7], it could be assumed that in subjects with high scores on the Private Self-Consciousness subscale, the increase of delta synchronization could be related to the high level of arousal when perceiving morphed images of faces that do not seem like themselves.

In the self-recognition condition, the Public Self-Consciousness/Social Anxiety subscale was positively related to an increase of theta rhythm. Based on the data on the relations between theta rhythm and attention and memory processes [8, 9], it could be assumed that subjects with a higher score on the Public Self-Consciousness/Social Anxiety subscale in the condition of recognizing themselves in a morphed image pay more attention to details, inconsistencies, and distortions of their own face.

Funding: The study was supported by the Russian Science Foundation (RSF) under Grant No. 22-15-00142.

References

1. Keenan J.P., Rubio J., Racioppi C., Johnson A., Barnacz A. The right hemisphere and the dark side of consciousness. *Cortex*. 2005;41(5):695-704. doi 10.1016/S0010-9452(08)70286-7
2. Gotlib I.H., Joormann J. Cognition and depression: current status and future directions. *Ann Rev Clin Psychol*. 2010;6:285-312. doi 10.1146/annurev.clinpsy.121208.131305
3. Mor N., Winquist J. Self-focused attention and negative affect: a meta-analysis. *Psychol Bull.* 2002;128(4):638-662. doi 10.1037/0033-2909.128.4.638
4. Fenigstein A. On the nature of public and private self-consciousness. *J Personality*. 1987;55(3):543-554. doi 10.1111/j.1467-6494.1987.tb00450.x
5. Bocharov A.V., Lebedkin D.A., Savostyanov A.N., Knyazev G.G. Adaptation of Russian versions of Self-focused attention scale and Self-consciousness scale. *Russ Psychol J*. 2023;20(3):97-115. doi 10.21702/rpj.2023.3.5 (in Russian)
6. Güntekin B., Başar E. Review of evoked and event-related delta responses in the human brain. *Int J Psychophysiol*. 2016;103:43-52. doi 10.1016/j.ijpsycho.2015.02.001
7. Knyazev G.G. EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neurosci Biobehav Rev*. 2012;36(1):677-695. doi 10.1016/j.neubiorev.2011.10.002
8. Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res Rev*. 1999;29(2):169-195. doi 10.1016/S0165-0173(98)00056-3
9. Knyazev G.G. Motivation emotion and their inhibitory control mirrored in brain oscillations. *Neurosci Biobehav Rev*. 2007;31:377-395. doi 10.1016/j.neubiorev.2006.10.004

doi 10.18699/bgrs2024-10-17

Dietary lecithin effects on the behavior, brain metabolism, mitochondrial and cellular structure of the neurons in mice

Boldyreva L.V.[1]*, Morozova M.V.[1], Medvedeva S.S.[1], Kozhevnikova E.N.[2], Morozova K.N.[3], Kiseleva E.V.[3]

[1] *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

[2] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *boldyrevalv@neuronm.ru*

Key words: behavior; soy lecithin; C57BL/6J laboratory mice; brain; metabolism; mitochondria; synapse

Motivation and Aim: Nowadays lecithin is widely used as a neuroprotective and hepatoprotective drug, a dietary supplement and is the often component of food production as an emulsifier. Therefore, the overall dose of lecithin in the modern human diet can be very high. Previously, we have shown that chronic intestinal inflammation in *Muc2*-knockout mice induces behavioral changes along with the substantial increases of the phospholipids content in intestinal epithelial cells, particularly, phosphatidylcholine, phosphatidylserine, and phosphatidic acid [1, 2]. It is known that phospholipids perform a whole range of molecular and cellular functions, and changes in their metabolism correlate with diseases and the course of chronic inflammatory processes [3]. Animals given the same phospholipid mixture with food have shown similar behavioral changes [4]. Also, our data demonstrate impaired mitochondrial function in intestinal cells of such animals [1, 5]. Soy lecithin consists up to 70 % of the similar phospholipids mixture: phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol and phosphatidic acid. Here we investigate the effects of soy lecithin administration on the behavioral patterns, brain cells and metabolism in laboratory mice.

Methods and Algorithms: Long-term feeding of C57BL/6 mice: pregnant females, starting from the third week of pregnancy, received standard food with the addition of soy lecithin (35 g per 1 kg, Solgar Inc. USA). The offspring also continued to receive lecithin in the same doses until behavioral testing. Short-term lecithin feeding: males received a soy lecithin with food (35 g per 1 kg) for four weeks before testing. Behavioral tests: assessment of anxiety, autistic-like behavior; social test; test for discrimination between social and non-social odor stimuli. Untargeted metabolomics analysis in mice brain samples: LC-MS/MS analysis of the metabolites composition (Waters Aquity UPLC/Q-Exactive HF-X (Thermo Scientific)) in brain samples was performed in Advanced Mass Spectrometry Core Facility of Skolkovo Institute of Science and Technology Metabolomics. Transmission electron microscopy (TEM) analysis was performed on ultrathin sections of hypothalamus, amygdala, and frontal cortex of mice brain at the Center for microscopy analysis of ICG SBRAS (FWNR-2022-0015).

Results: As a result of experiments on long-term and short-term feeding of healthy C57BL/6 laboratory mice with soy lecithin, significant deviations in social behavioral patterns were revealed. In particular, animals did not distinguish between the odor of a female and a male, while their discrimination of non-social odors was normal. At the

same time, in the test with two intruders (female and male), animals after lecithin feeding did not demonstrate normal preference towards the female. In addition, perinatal, rather than short-term feeding of lecithin resulted in changes in stereotypic behavior. These behavioral changes were accompanied by changes in a number of metabolites levels identified by LC-MS analysis in the brain of experimental group of animals. TEM analysis revealed substantially altered structure of mitochondria in hypothalamus and amygdala regions of brain in animals with long-termly administered soy lecithin: the density of the matrix increases, and internal structures degrade forming empty space inside the mitochondria. In addition, synaptic vesicles were found to be of irregular size and shape, and the synaptic cleft formed a wider than normal brush-like structure. In experimental animals, high rate of the apoptosis of neurons and glial cells, as well as cellular membranes degradation, was more often observed.

Conclusion: Our results indicate a significant effect of lecithin supplementation on brain function, metabolism and cellular structure of neurons in laboratory mice. Further studies of the effects of the lecithin and dietary phospholipids, as well as their metabolic regulatory capacity on the brain functions regulation of may provide a new level in understanding the role of metabolic compounds in the gut-brain axis.

Funding: The study was supported by RSF No. 23-25-00417 (https://rscf.ru/project/23-25-00417/).

References

1. Morozova M.V., Borisova M.A., Snytnikova O.A. et al. Colitis-associated intestinal microbiota regulates brain glycine and host behavior in mice. *Sci Rep*. 2022;12(1):16345. doi 10.1038/s41598-022-19219-z
2. Borisova M.A., Achasova K.M., Morozova K.N. et al. Mucin-2 knockout is a model of intercellular junction defects, mitochondrial damage and ATP depletion in the intestinal epithelium. *Sci Rep*. 2020;10(1):21135. doi 10.1038/s41598-020-78141-4
3. Boldyreva L.V., Morozova M.V., Saydakova S.S., Kozhevnikova E.N. Fat of the Gut: Epithelial Phospholipids in Inflammatory Bowel Diseases. *Int J Mol Sci*. 2021;22(21):11682. doi 10.3390/ijms222111682
4. Boldyreva L.V., Morozova M.V., Pavlov K.S. et al. Effect of Dietary Phospholipid on the Behavior in C57BL/6J Mice. *J Evol Biochem Physiol*. 2024;60:409-419. doi 10.1134/S0022093024010319
5. Saydakova S., Morozova K., Snytnikova O. et al. The Effect of Dietary Phospholipids on the Ultrastructure and Function of Intestinal Epithelial Cells. *Int J Mol Sci*. 2023;24(2):1788. doi 10.3390/ijms24021788

doi 10.18699/bgrs2024-10-18

Longitudinal physiological and proteomic profiling under psychological assessments: a pilot study

Gorbunov K.[1]*, Kurilova O.[1], Engin D.[1], Makhnach A.[2]

[1] *Research Institute for Systems Biology and Medicine (RISBM), Moscow, Russia*

[2] *Institute of Psychology of Russian Academy of Sciences, Moscow, Russia*

** k.gorbunov@sysbiomed.ru*

Key words: proteomics; repeated measures; physiology; psychological assessments; survey

Motivation and Aim: Human behaviour is an integral function of the whole organism. Differences in the human proteome may reflect differences between human states [1–4]. We have attempted to identify patterns of dynamic changes in the human proteome over several days of observation, in connection with general physiological indicators of human adaptive changes and individual psychological characteristics. Previous attempts have been made to identify intergroup and individual differences in proteomic profiles of people due to their habits, age, sex, and various pathologies. Our study seeks to identify individual differences in the patterns of temporal changes in the human proteome. At the same time, the levels of part of the proteins remain unchanged due to homeostatic reasons, and part reflects current changes in the human body.

Methods and Algorithms: A longitudinal study was conducted, during which 20 patients were examined with respect to their proteomic states using 228 proteins and physiological conditions (drawn on FCI (Functional Change Index, R.M. Baevskij, 1997 [5, 6]) and Kerdo scores) over three days at four time points during the day (with two-hour time lag). Additionally, participants were classified into groups using Pavlovian Temperament Survey (PTS) and The Big Five Inventory (BFI). We used GEE and LMM to assess proteomic values and FCI scores as well as their respective within-subject correlation; repeated measures ANOVA to evaluate FCI dynamics over time; correlation analysis to estimate the relationship between viability score and proteins as well as viability score with one's level of arousal using nonparametric tests with permutation.

Results: It was concluded that dynamics of FCI scores over time is significantly different for two contrasting psychological groups. Two proteins were found to have significant within-subject correlation coefficients with respect to FCI scores and were significant using GEE and LMM after adjustment for psychological groupings; seven proteins were obtained as crucial when evaluating relationship between viability scores and proteomic profiles.

Conclusion: We hypothesize that dynamics of a proteomic profile reflects physiological and psychological conditions of an individual, we intend to continue our research for further evaluation of the given proposition.

Funding: The study is supported by ГЗ №141-03-2023-/05 since 31.01.2023 (ГЗ -2023 ЕГСУ НИОКТР 122030900062-5).

References

1 Carrasco-Zanini J. et al. Proteomic signatures for identification of impaired glucose tolerance. *Nat Med.* 2022;28(11):2293-2300. doi 10.1038/s41591-022-02055-z

2 Nimer R.M., Alfaqih M.A., Shehabat E.R., Mujammami M., Abdel Rahman A.M. Label-free quantitative proteomics analysis for type 2 diabetes mellitus early diagnostic marker discovery using

data-independent acquisition mass spectrometry (DIA-MS). *Sci Rep*. 2023;13(1):20880. doi 10.1038/s41598-023-48185-3
3 Oh H. S.-H. et al. Organ aging signatures in the Plasma Proteome Track Health and disease. *Nature.* 2023;624(7990):164-172
4 Williams S.A. et al. Plasma protein patterns as comprehensive indicators of health. *Nat Med.* 2019;25(12):1851-1857
5 Baevskij R.M. Predicting conditions on the verge of normality and pathology. Moscow: Medicina, 1979 (in Russian)
6 Baevskij R.M., Berseneva A.P. Assessment of the adaptive capabilities of the body and the risk of developing diseases. Moscow: Medicina, 1997 (in Russian)

doi 10.18699/bgrs2024-10-19

Toward human-like intelligence in artificial systems: the feeling of knowing

Kralik J.D.*
Department of Brain and Cognitive Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
** jerald.kralik@gmail.com*

Key words: artificial general intelligence; affect; subjective experience; neuroscience; cell signalling

Motivation and Aim: Human intelligence includes a sense of knowing. Patterns and probabilities are important (and used by the brain), but they result from something more to the point: the reasons why they are observed, what they result from. And human-like knowing cares about these underlying reasons. That is, it is making sense of things (see [1]). And to make sense, this theoretical analysis argues, requires not only determining such things as *what* is happening (e. g., what or who is involved), *why* (the causal agents), *how* (mechanism of change), the *expected outcome*, and *evaluating* it (typically with respect to valence of good or bad), but also the *feeling* one has when they believe they have understood the situation properly. The *first aim* of the current report is to clarify this 'feeling of knowing' and to show how an evolutionary analysis reveals its significance for human intelligence. The *second aim* is then to circumscribe its possible origins in humans, and thereby provide a framework for more detailed computational and biological work to (a) determine exactly how it emerges and influences human intelligence, and in turn, (b) model it for integration in artificial systems.
Methods: The theoretical analysis takes as premise that the *feeling of knowing* is of the same essential character as general affect, emotion, care, and even subjective awareness underpinning consciousness. Based on this, I attempt to examine and outline all possible sources of the feeling of knowing, highlighting the main findings in affective neuroscience (such as structures in the limbic system), but then zooming out with an evolutionary and biological perspective: considering cells of single- (both prokaryote, like bacteria, and eukaryote, like algae) and multicellular organisms; systems of eukaryote cells in multicellular organisms such as ourselves (pointing to intra and extra cellular coordination and signaling); the uniqueness of neural signals of the nervous system; and the scale and organization (including hierarchy and loops) achieved via layers of signals reflecting *need/affect/care* (for survival and reproduction and beyond), that give rise to the sense of care and feeling in other animals, and ultimately in true human understanding.
Results: Four key considerations each led to critical insights. First, to address the question of how significant the 'feeling of knowing' is for higher intelligence, we look to the relevant research in psychology and neuroscience. The literature thus far, however, provides mixed results on the extent *cognition*, that is, knowledge and thinking, can be separated from affect (and feeling). An evolutionary approach can provide insight. And to do this requires considering the origins of the most foundational *sensing-cognition-action cycle* in the history of life: that is, with individual (prokaryotic and then eukaryotic) cells, and then the progression to multicellular organisms. Put differently, we can examine which came first – sensing, action, cognition, or need – and then how

these general components interact. **Insight one:** The most foundational factor is *need* (for resources and homeostasis), followed by *need-driven action* (to attain what is needed), which then requires *sensory guidance*, with continued elaboration of sensory, action, and need processing via *cognition*. And importantly, all elements of this fundamental cycle (need, action, sensation, cognition) appear tightly coupled (i. e., recurrent rather than simply sequential), with for example, action anticipation influencing sensory processing.

The next question is whether the need-sensation-cognition-action cycle components remain tightly coupled as one ascends from simpler (e. g., cells to nervous systems, brainstem to hypothalamus) to more complex regions (sensorimotor cortex to the highest association areas), especially in humans. **Insight two:** tight coupling should and does appear to occur at all levels of the human brain. Focusing on need or affect specifically, although some affective neuroscientists have provided evidence for affective influences at the highest levels of the human brain and thinking, the extent of the influence remains underappreciated [1, 2]. But as suggested by evolutionary psychology and biology, and by the specialized nuclei in lower and mid brain structures (e. g., hypothalamus), the human brain appears to be organized around the fundamental *needs* of life (e. g., resource attainment, especially fluid and food, defense, mating, etc.), with affective-emotional instinct and knowledge priors producing specializations even at the highest levels of cortex: with the clearest higher-level example being the so-called social brain [1].

Is the 'feeling of knowing', then, essentially synonymous with affect and reward, providing the drive (wanting) and payoff (liking) for goal attainment? **Insight three**: It is, of course, more than this. Nonetheless, it is clear that for all aspects of intelligence – such as knowledge seeking, attaining and utilizing (e. g., in reasoning) – affective-emotional-subjective influences, elaborated into higher-level versions, are involved: such as the sense of delight, empowerment, prestige, and satisfaction with problems solved, knowledge gaps filled, discoveries made [3]. And a strong implication of the current analysis suggests that without such sensations, such as the relief and satisfaction one feels when a correct hypothesis turns confusion to clarity and broader understanding, true knowing and belief remain incomplete.

Finally, where and how does this 'feeling of knowing' originate in the brain – how does it emerge? The answer indeed remains out-of-reach, and so requires (a) a description of what is currently known; and (b) a roadmap leading to a comprehensive understanding. **Insight four:** there are many possible sources producing emergent subjective experience, which together construct such a framework and roadmap. For humans, and vertebrates more generally, *limbic* brain structures have been especially implicated [2]. More specifically, there is strong, compelling evidence for the conversion of neural signals representing sensory information (e. g., higher firing rate with larger food quantity: large chocolate bar!) to subjective experience (e. g., lower firing rate with high food quantity due to satiation: no more chocolate!) in these limbic structures. Prominent examples include the orbitofrontal cortex (taste), the nucleus accumbens of the subcortical basal ganglia (wanting and liking), and the insula cortex (pain) [1]. But zooming in remains unclear. For example, for *reward*, *wanting* and *liking* components have been tied to dopamine and endogenous opioid infusion (from neurons in the midbrain) into the nucleus accumbens, respectively [4, 5]. And so neurochemistry is critical. But what is happening? The neurochemicals open and close channels to excite and inhibit various neural populations. But how subjective experience derives or emerges from the influence of such infusions is unknown. Zooming out suggests a

potentially wider influence of 'cortical, basal ganglia, thalamus, back to cortex' loop structures (let alone the bottom-up influence of the hypothalamus and midbrain as well) [1, 2, 6, 7]. And so the subjective experience of the feeling of knowing potentially results from the recurrence of these loops and/or from other aspects of the greater complexity of this larger system. Critically, however, a great deal of evidence suggests that other organisms at much simpler levels of nervous system complexity also show signs of affect and subjective experience [8]. And so comparative work should also help. Indeed, returning to the prior evolutionary examination, there is *need* at all levels: i. e., inside every cell, with organelles and other cellular structures sending *signals* to communicate this need. For any given cell, let alone the multicellularity of even very simple organisms, tens of thousands to millions of signals lead to activation of homeostasis systems and affect. Whether the critical factor for emergence results from the specific character of the neural signal (together with activating neurochemistry) without significant influence from the massive *scale* (and hierarchical organization) of nonneural cellular signals remains an open question. Examination on all fronts is warranted. This importantly includes more computational work, to help clarify all roles and continue circumscribing and triangulating the source(s).

Conclusion: One of the most vexing and intractable elements of human intelligence is in fact an affective-type component: not only the desire to know, but the sensation – or feeling – of knowing; which is believed and argued here to be required for true human intelligence, and therefore for comparable artificial systems. My theoretical analysis, together with other theoretical and empirical work, circumscribes a great deal of possible sources leading to the emergence of this subjective experience, ranging from the sheer scale and extraordinary computational complexity in layered and recurrent architectures, to potential specific aspects of the actual chemistry – though the latter possibility remains underspecified, as it is unclear what of the chemistry could not be artificially reproduced. Although the hardware in biological and most artificial systems (like computers) are different, they nevertheless share fundamental-force underpinnings (electromagnetic, strong, weak). Nevertheless, the insights obtained here, though daunting, still suggest significant progress can be made, with concerted efforts along multiple lines of inquiry, toward understanding and thereby ultimately reproducing 'feeling' for true human-like knowing in artificial systems.

References

1. Gazzaniga M.S., Ivry R.B., Mangun G.R. Cognitive neuroscience: the biology of the mind. WW Norton & Company, 2018
2. LeDoux J.E. Emotion Circuits in the Brain. *Ann Rev Neurosci*. 2003;23(1):155-184
3. Loewenstein G. The psychology of curiosity: A review and reinterpretation. *Psychol Bull*. 1994;116(1):75-98. doi 10.1037/0033-2909.116.1.75
4. Morales I., Berridge K.C. 'Liking' and 'wanting' in eating and food reward: Brain mechanisms and clinical implications. *Physiol Behav*. 2020;227:113152. doi 10.1016/j.physbeh.2020.113152
5. Jang H., Jung K., Jeong J., Park S.K., Kralik J.D., Jeong J. Nucleus accumbens shell moderates preference bias during voluntary choice behavior. *Social Cognit Affective Neurosci*. 2017;12(9):1428-1436. doi 10.1093/scan/nsx072
6. Kralik J.D. Architectural Design of Mind & Brain from an Evolutionary Perspective. AAAI Fall Symposium: Standard Model of Mind, 2017
7. Kralik J.D. Toward a Comprehensive List of Necessary Abilities for Human Intelligence, Part 1: Constructing Knowledge. In: Proceedings of the 15th International Conference on Artificial General Intelligence (AGI 2022). 2022;271-281. doi 10.1007/978-3-031-19907-3_25
8. Koch C. The Quest for Consciousness: A Neurobiological Approach. Roberts & Co, 2004

doi 10.18699/bgrs2024-10-20

Development of a hybrid tool system for semi-automatic ontology creation

Lebedev M.[1]*, Savostyanov A.[1, 2, 3]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

* *ml220501@gmail.com*

Key words: ontologies; automation; tool; automatic ontology creation; mental illnesses; data structurization

Motivation and Aim: Nowadays, the excess of the information on the web hinders research and finding relevant information. For example, when searching for papers that include keywords “Depression” and “Genes” there are approximately 12,000 results. And when searching using keywords “Proteins” and “Neurotransmitter systems” about 137,000 results are given. That means that finding data connections is difficult if not impossible.

Ontologies aim to fix this problem by providing data systematization for different fields. For example, ontologies are widely used in medicine, one of them being ANDDigest that helps find genes associated with certain organisms and illnesses, along with a lot of other information [1–3]. However, creating anthologies is a time-consuming and difficult process that requires a lot of specific expert knowledge. This led to the development of tools that automate certain parts of ontology creation. The problem with automatic approach is that it is usually based on deduction, rather than induction, which can provide false information.

These negatives of both approaches (manual and automated) can be solved by creating a hybrid system utilizing automation algorithm and expert knowledge.

Methods and Algorithms: In this work a 4-level ontology system will be used [4]. The four levels are: ontological knowledge, general theoretical knowledge, empirical knowledge, probabilistic knowledge. Automated ontology creation steps correlate with ontology system levels: term extraction for ontological knowledge, link extraction for empirical knowledge and axiomatics extraction for theoretical knowledge.

After studying existing ontology creation tools, it is clear that there are two approaches: linguistic and statistical [5, 6].

Linguistic tools are most commonly used for text preprocessing, as well as parsing part-of-speech and syntactic information. Later this information can be used for term and link extraction.

Statistical tools are also used for term and link extraction, however they use very different principals. Using frequency and statistic information lets these tools become language independent.

Results: At this stage, a basis for tool creation has been developed. Firstly, it includes the tool’s structure: it will consist of several modules (axiomatics creation, text parsing, term extraction and link extraction) with the first one being an expert module, others being automatic and the latter two being expert-verified. Secondly, the requirements for the tool have been formulated: Technology stack – Python will be used for the tool and one of the more universal ontology systems will be used for automatic modules

(OntoGain or text2onto) as this provides the ability to use both statistical and linguistic algorithms without having to switch ecosystems; Integration – the tool must be compatible with modern existing ontology editors (like Protege) by utilizing commonly used formats (for example, omn or rdf); Simplicity – the tool must be usable by a person not fluent in computer science, which implies have an intuitive user interface; Discreteness – different modules need to be able to function on their own to save time and make testing and ontology creation more effective. Lastly, the list of classes that will be used in a final test ontology has been created. It was done by hand using Protege in collaboration with a field expert, however in the future, a special module will be developed to simplify this process.

Conclusion: Currently, all the necessary preparations for development have been made and the next step will involve integrating automation algorithms in a tool and testing it on scientific papers on mental illnesses and their studying and causes.

Funding: The study is supported within the budget project of the ICG SB RAS No. FWNR-2022-0020.).

References

1. Ivanov R., Kazantsev F., Zavarzin E. et al. ICBrainDB: An Integrated Database for Finding Associations between Genetic Factors and EEG Markers of Depressive Disorders. *J Pers Med*. 2022;12(1):53. doi 10.3390/jpm12010053
2. Ivanov R., Zamyatin V., Klimenko A., Matushkin Y., Savostyanov A., Lashin S. Reconstruction and Analysis of Gene Networks of Human Neurotransmitter Systems Reveal Genes with Contentious Manifestation for Anxiety, Depression, and Intellectual Disabilities. *Genes (Basel)*. 2019;10(9):699. doi 10.3390/genes10090699
3. Ivanisenko T.V., Saik O.V., Demenkov P.S. et al. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics*. 2020;21(Suppl. 11):228. doi 10.1186/s12859-020-03557-8
4. Naydanov C.A., Palchunov D.E., Sazonova P.A. Model-theoretic methods of integration of knowledge extracted from medical documents. *Vestnik NSU. Series: Information Technologies*. 2015;13(3):29-41 (in Russian)
5. Zagorulko Yu.A., Sidorova E.A., Zagorulko G.B., Akhmadeeva I.R., Sery A.S. Automation of the development of ontologies of scientific subject domains based on ontology design patterns. *Ontologiya proektirovaniya*. 2021;11(4):500-520. doi 10.18287/2223-9537-2021-11-4-500-520 (in Russian)
6. Asim M.N., Wasim M., Khan M.U.G., Mahmood W., Abbasi H.M. A survey of ontology learning techniques and applications. *Database (Oxford)*. 2018;2018:bay101. doi 10.1093/database/bay101

doi 10.18699/bgrs2024-10-21

Perception of self- and other-referential sentences with varying emotional content: a neurolinguistic study

Lebedkin D.A.[3]*, Saprygin A.E.[1, 2], Vergunov E.G.[1, 3], Savostyanov A.N.[1, 2, 3]

[1] *Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *d.lebedkin@g.nsu.ru*

Key words: neuropsychology; neurolinguistics; Self; emotion processing; syntax processing; EEG

Motivation and Aim: Implicit neurophysiological markers of psychological states and characteristics are widely utilized in the field of experimental psychology and psychiatry as they allow to study measurable and minimally biased reactions. In this exploratory study we aim to find new markers of language emotional processing and self-reference using spatiotemporal permutational test with clustering for electroencephalography (EEG) event-related potentials (ERP) data.

Methods and Algorithms: EEG datasets of sixty-four participants (34 females) aged between 18 and 52 ($\mu = 23.1$, SD = 7.2) were selected for this study. The datasets were acquired during a series of our previous behavioral experiments described in more detail in [1]. In short, a 128-channel EEG recording was made during which subjects were shown Russian sentences containing grammar errors and asked whether a sentence was correct or not. Unbeknownst to the participants, these sentences also described situations of anxiety or aggression and were referencing self (the reader) or other people (or were neutral/referenced an inanimate object). Two hundred sentences in total were shown, each from one of 10 categories based on emotional content, reference and grammatical correctness. EEG datasets were preprocessed automatically, using custom Python scripts and MNE, SciPy, NumPy, Pandas libraries. The pipeline for preprocessing was as follows: (1) bandpass filtering 0.1–100 Hz, notch filtering for 50 Hz; (2) detection and interpolation of noisy channels; (3) REST re-reference [2]; (4) ICA for noise reduction; (5) epoching with 0.5 s pre-stimulus and 1 s post-stimulus; (6) epoch-wise interpolation of noisy channels; (7) epoch averaging. Preprocessed data were then used to form a matrix of size $N_{subjects}*N_{channels}*N_{timepoints}$ (67*128*1500) to be used in a spatiotemporal permutation test with clustering as implemented in MNE Python [3–5]. We used the following set of parameters for the test: $N_{permutations}$=10.000; p-value for spatial cluster formation ≤ 0.001.

Results: Spatiotemporal permutation test with clustering for all conditions showed 2 significant clusters ($p_1 = 0.0015$, $p_2 = 0.0008$) of varying duration (50 ms and 100 ms) and onset (560 ms and 660 ms), but similar spatial placement (medial central electrodes) as shown in the Fig. 1 and 2. Fig. 3 and 4 show results for comparison of conditions anxiety/aggression/neutral (Fig. 3) and self/others/neutral (Fig. 4) and each have a significant cluster ($p_3 = 0.0011$, $p_4 = 0.0016$) of approximately the same duration (250 and 260 ms) and onset (550 ms), and spatial placement (medial central electrodes).

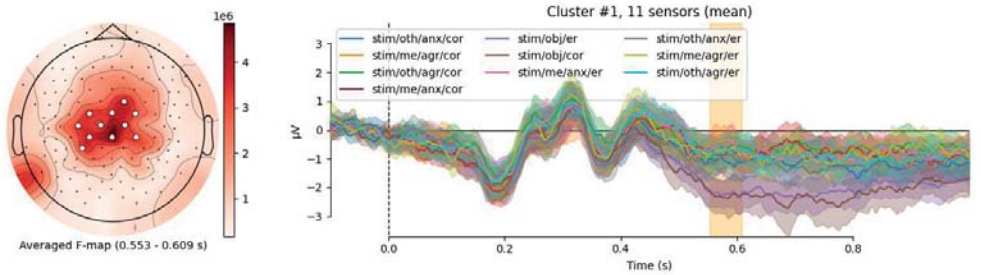


Fig. 1. Permutation test results for all conditions, cluster #1

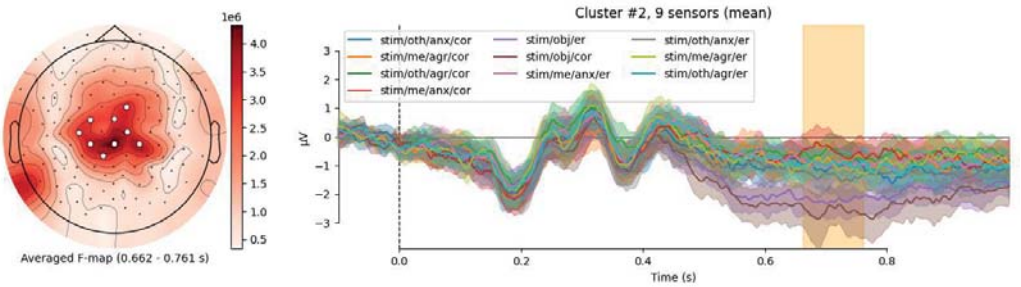


Fig. 2. Permutation test results for all conditions, cluster #2

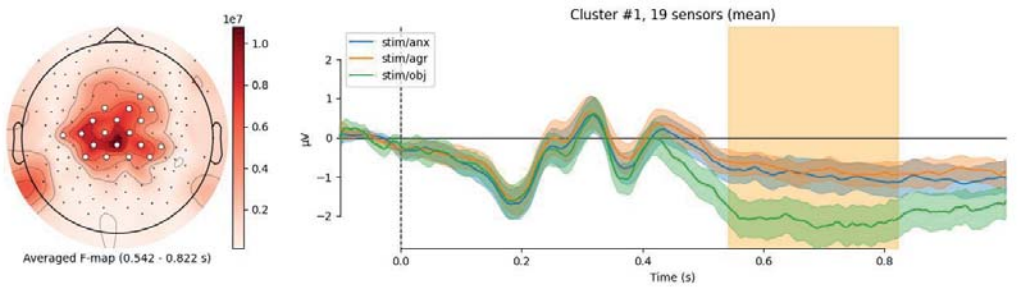


Fig. 3. Permutation test results for conditions anxiety vs. aggression vs. objective

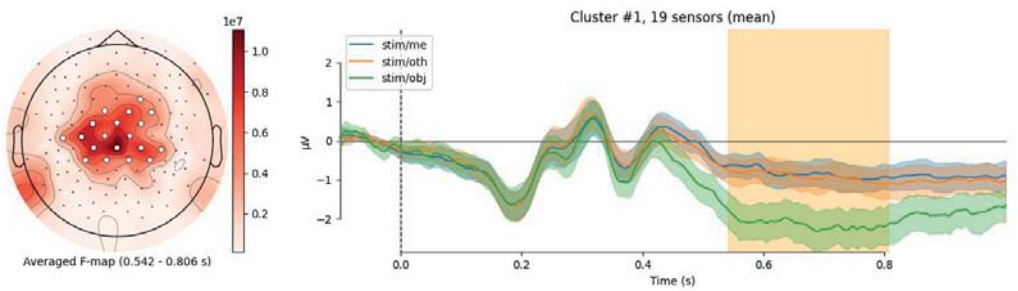


Fig. 4. Permutation test results for conditions self vs. other vs. objective

Conclusion: Our results point out that the late ERP time period (550–800 ms) may be important for linguistic emotional processing, which is also supported by the literature [6]. Neutral sentences are shown to elicit the most negative response in the later time period, which may have implications for their usage as “baseline” stimuli. Spatial characteristics of found clusters draw our attention to medial frontoparietal cortex regions.

Funding: The study is supported by the Russian Science Foundation (project No. 22-15-00142).

References

1. Savostyanov A. et al. The Effect of Meditation on Comprehension of Statements About One-Self and Others: A Pilot ERP and Behavioral Study. *Front Hum Neurosci*. 2020;13:437. doi 10.3389/fnhum.2019.00437
2. Yao D. A method to standardize a reference of scalp EEG recordings to a point at infinity. *Physiol Meas.* 2001;22(4):693-711. doi 10.1088/0967-3334/22/4/305
3. Gramfort A. et al. MEG and EEG data analysis with MNE-Python. *Front Neurosci*. 2023;7:267. doi 10.3389/fnins.2013.00267
4. Maris E., Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods*. 2007;164(1):177-190. doi 10.1016/j.jneumeth.2007.03.024
5. Smith S.M., Nichols T.E. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*. 2009.;44(1):83-98. doi 10.1016/j.neuroimage.2008.03.061
6. Weinberg A., Correa K.A., Stevens E.S., Shankman S.A. The emotion-elicited late positive potential is stable across five testing sessions. *Psychophysiology*. 2021;58(11):e13904. doi 10.1111/psyp.13904

doi 10.18699/bgrs2024-10-22

Changes in salivary oxytocin levels, cardiac vagal tone, and functional connectivity in the brain: effects of voluntary breathing regulation

Lee S.-K.[1, 2], Lee J.-D.[2], Amunts K.[3, 4], Bludau S.[3], Huang A.C.W.[5], Tsai A.C.[2], Liou M. [2]*

[1] *Taiwan International Graduate Program in Interdisciplinary Neuroscience, National Cheng-Kung University & Academia Sinica, Tainan 701, Taiwan*

[2] *Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan*

[3] *Institute of Neuroscience and Medicine, Structural and Functional Organization of the Brain (INM-1), Research Centre Jülich, Jülich, Germany*

[4] *C. and O. Vogt Institute for Brain Research, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany*

[5] *Department of Psychology, Fo-Guang University, Yilan, Taiwan*

**mliou@stat.sinica.edu.tw*

Key words: Breathing regulation; HRV indices; resting-state fMRI; salivary oxytocin concentrations; stress modulation

Motivation and Aim: Breathing is one of the most fundamental and partially controllable physiological behaviors in daily life. Breathing regulation has been identified as a clinical method to temporarily relieve anxiety and depressive symptoms. However, It is still an ongoing debate in the scientific community regarding the functional circuits and voluntary breathing regulation within the brain.

Methods: The study was divided into two core sections: the first section explored changes in functional magnetic resonance imaging (fMRI) time courses during rapid breathing and breath-holding cycles, as well as potential respiratory-related connectivity networks in the brain. The second section recruited two groups of participants – the experimental group underwent respiratory training exercise, and the sham group played piano keyboard games – to examine group differences within resting-state fMRI connectivity networks and on their salivary hormone concentrations.

Results: The first experiment found that there were at least two connectivity networks highly associated with the interactions between breathing regulation and ANS. Notably, the ventral medial prefrontal cortex exhibited fMRI time courses that showed inverse patterns to those observed in the dysgranular structure in the insula cortex, amygdala, and premotor cortex. The second experiment revealed significant increases in salivary oxytocin levels in participants after breathing exercise compared with those playing piano games. Additionally, the experimental group demonstrated significantly weaker inter-hemispheric connectivity in the hippocampus and its associated motor areas as well as weaker connectivity between the hippocampus and the salience network.

Conclusion: This study identified brain connectivity networks associated with breathing regulation and their associations with salivary oxytocin levels. The results are expected to provide a neurophysiological basis and connectivity evidence for the clinical application of respiratory practices in emotion regulation, and for enriching the scientific foundation of psychiatric therapy.

Funding: This work was supported by the National Science Council, Taiwan, under Grants NSC110-2410-H-001-046, and 112-2410-H-001-068.

References

1 Amunts K., Mohlberg H., Bludau S., Zilles K. Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science*. 2020;369(6506):988-992

2 Carter C.S., Kenkel W.M., MacLean E.L., Wilson S.R., Perkeybile A.M., Yee J.R., Ferris C.F., Nazarloo H.P., Porges S.W., Davis J.M. Is oxytocin "nature's medicine"? *Pharmacol Rev*. 2020;72(4):829-861

3 Chang S.-D., Kuo P.-C., Zilles K., Duong T.Q., Eickhoff S.B., Huang A.C., Tsai A.C., Cheng P.E., Liou M. Brain Reactions to Opening and Closing the Eyes: Salivary Cortisol and Functional Connectivity. *Brain Topogr*. 2022;35(4):375-397

4 Gianaros P.J., Wager T.D. Brain-Body Pathways Linking Psychological Stress and Physical Health. *Curr Directions Psychol Sci.* 2015;24(4):313-321

5 Magnon V., Dutheil F., Vallet G.T. Benefits from one session of deep and slow breathing on vagal tone and anxiety in young and older adults. *Sci Rep*. 2021;11(1):19267

6 Rokicki J., Kaufmann T., De Lange A.-M.G., van der Meer D., Bahrami S., Sartorius A.M., Haukvik U.K., Steen N.E., Schwarz E., Stein D.J. Oxytocin receptor expression patterns in the human brain across development. *Neuropsychopharmacology*. 2022;47(8):1550-1560

7 Jiang Y., Travagli R.A. Hypothalamic – vagal oxytocinergic neurocircuitry modulates gastric emptying and motility following stress. *J Physiol*. 2020;598(21):4941-4955

11 Симпозиум «Фундаментальные генетические/клеточные системы/процессы: компьютерные и экспериментальные подходы»

Symposium “Fundamental genetic/cellular systems/processes: computational and experimental approaches”

doi 10.18699/bgrs2024-11.1-01

Природные полиморфизмы ферментов эксцизионной репарации оснований ДНК как фактор эффективности удаления повреждений

Алексеева И.В., Кладова О.А., Кузнецов Н.А.*

Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** nikita.kuznetsov@niboch.nsc.ru*

Ключевые слова: репарация ДНК; однонуклеотидный полиморфизм; ферментативная активность

Мотивация и цель: Стабильность генетической информации в живой клетке – ключ к ее нормальному функционированию. На сегодняшний день не остается сомнений, что повреждения ДНК играют ключевую роль в возникновении различных заболеваний, в том числе онкологических. Процесс эксцизионной репарации оснований (BER) включает действие множества ферментов и отвечает за удаление из ДНК огромного числа повреждений. Необходимо отметить, что уменьшение общего уровня репарационной активности у человека может быть вызвано природными однонуклеотидными полиморфизмами (SNPs) в генах ферментов репарации ДНК. Действительно, SNPs широко распространены в популяции людей, а взаимосвязь между SNP и этиологией некоторых заболеваний человека остается открытой областью исследований. При этом один из актуальных вопросов заключается в выяснении роли SNPs в предрасположенности к развитию онкологических и нейродегенеративных заболеваний. Следует также отметить, что SNP-ассоциированные замены аминокислотных остатков ферментов репарации могут приводить как к потере функциональных свойств отдельных ферментов, так и к нарушению белок-белковой координации участников процесса репарации. Целью данной работы являлся систематический анализ природных полиморфных вариантов ряда ферментов эксцизионной репарации оснований ДНК человека, а именно ДНК-гликозилаз SMUG1 и MBD4, АР-эндонуклеазы APE1 и ДНК-полимеразы β.
Методы и алгоритмы: С использованием базы данных dbSNP NCBI (http://www.ncbi.nlm.nih.gov/SNP/) было установлено, что в генах ДНК-гликозилаз SMUG1 и MBD4, АР-эндонуклеазы APE1 и ДНК-полимеразы β зарегистрированы тысячи SNPs. Методами биоинформационного анализа показано, что некоторые из них могут оказывать влияние на эффективность действия ферментов. Проведено молекулярное моделирование ряда выбранных мутантных вариантов ферментов для оценки возможного влияния замен аминокислотных остатков на каталитическую активность и способность формировать белок-белковые комплексы. Для верификации полученных результатов проведен экспериментальный анализ ферментативных свойств ряда рекомбинантных вариантов, который выявил существенное влияние отобранных природных замен как на каталитическую активность ферментов, так и на их координацию с другими участниками процесса репарации.
Результаты: В настоящий момент существует большое количество биоинформатических подходов, позволяющих предсказывать влияние

однонуклеотидных (SNP) замен на функцию белка. Нами был проведен анализ известных полиморфизмов ДНК-гликозилаз SMUG1 и MBD4, АР-эндонуклеазы APE1 и ДНК-полимеразы β человека с целью выявления полиморфных вариантов, способных оказывать повреждающее воздействие на функционирование данных ферментов. Информация об известных однонуклеотидных полиморфизмах была получена из базы данных NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp). Особое внимание было уделено полиморфизмам, которые (1) приводили к смене класса аминокислотного остатка, (2) аминокислотные остатки располагались в функционально важных участках белковой глобулы и/или (3) данные SNP встречались в опухолях. Эти критерии позволили отобрать ряд SNP-ассоциированных замен в четырех ферментах для их дальнейшего анализа.

Анализ влияния природных полиморфизмов на ферментативную активность APE1 (R221C, N222H, R237A, G241R, M270T, R274Q, P311S) и ДНК-гликозилаз SMUG1 (G90C, P240H, N244S и N248Y) и MBD4 (S470L, G507S, R512W и H557D) показал, что, если для некоторых вариантов ДНК-гликозилаз обнаружена практически полная потеря ферментативной активности (SMUG1 G90C и P240H, MBD4cat S470L, G507S и R512W), то в случае APE1, по-видимому, активность фермента не может быть ниже порогового значения, необходимого для поддержания минимального уровня репарационной активности, достаточного для выживания организма. Такое различие в результатах может быть связано с тем, что в случае ДНК-гликозилаз их субстратная специфичность может дублироваться другими ферментами этого класса, в то время как снижение активности APE1 приводит к крайне негативным для организма последствиям.

При определении активности мутантных вариантов POLβ в трансферазной реакции установлено, что все исследованные мутантные формы (L19P, G66R, G118V, R149I, E154A, G189D, M236T, R254I, G274R, G290C, R333W) обладали более медленной способностью заполнять 1 нт бреши в ДНК, чем фермент дикого типа. Следует отметить, что для мутантных форм, несущих замены G274R и R333W, не удалось зарегистрировать накопление продукта реакции, что указывает на значительную степень инактивации ферментов, при возникновении замен аминокислотных остатков в положении G274 и R333.

Эффективность координированных взаимодействий между APE1 и ферментами, выполняющими свои функции как до (ДНК-гликозилазы: OGG1, UNG2, AAG), так и после APE1 (POLβ) в процессе BER, а также с белковыми факторами (XRCC1 и PCNA) была определена по способности белков-эффекторов оказывать влияние на активность APE1 дикого типа и ряд ее природных полиморфных вариантов. Показано, что аминокислотные остатки Arg221 и Arg237, располагающиеся на поверхности белковой глобулы, важны для регуляции активности APE1 другими белками репарационного комплекса, но при этом оказывают различное влияние на активность APE1. Тогда как Asn222, по-видимому, принимает активное участие в контактах не со всеми протестированными белками. Так, например, белки-эффекторы OGG1, UNG2 и AAG практически не ведут к изменению активности этого SNP-варианта. Внесение аминокислотных замен в ДНК-связывающий центр APE1 приводит к нарушению координированных взаимодействий с другими участниками BER. В пользу этого вывода свидетельствует факт, что оба SNP варианта APE1, M270T и R274Q, наименее чувствительны к стимуляции протестированными белками-эффекторами по сравнению с другими полиморфными вариантами APE1.

Появление дополнительного положительного заряда в белковой глобуле в случае замены G241R приводит к снижению эффекта стимуляции. На вариант P311S наибольший стимулирующий эффект оказывают AAG и PCNA. Кроме того, для дикого типа урацил-ДНК-гликозилаз SMUG1 и MBD4^{cat}, а также их полиморфных вариантов SMUG1 N248Y и MBD4^{cat} H557D показан стимулирующий эффект каталитически неактивной формы АР-эндонуклеазы APE1 D210N.
Определение влияния белков-эффекторов (APE1, AAG, MBD4, NEIL1, NTHL1, OGG1 и TDG) на накопление продукта трансферазной реакции POLβ выявило, что добавление фермента APE1 в реакцию либо полностью ингибировало, либо замедляло (WT POLβ, G66R, R149I) накопление продукта для всех исследованных вариантов POLβ. Кроме того, для вариантов POLβ, обладающих низкой константой полимеризации (например, L19P, G118V, G189D), было зафиксировано образование продуктов экзонуклеазной активности APE1. ДНК-гликозилазы AAG, NEIL1 и NTHL1 снижали активность исследованных вариантов POLβ более чем на 20 %. Присутствие в реакции ферментов MBD4^{cat} и TDG практически не влияло на накопление продукта реакции отдельных вариантов POLβ, тогда как добавление фермента OGG1 приводило к небольшому снижению накопления продукта реакции, однако не ингибировало активность вариантов POLβ по сравнению с ферментом APE1.
Выводы: За время выполнения проекта проведен экспериментальный анализ 26 природных вариантов SMUG1, MBD4, APE1 и POLβ, который выявил существенное влияние ряда природных замен как на каталитическую активность ферментов, так и на их координацию с другими участниками процесса репарации. Полученные данные подтверждают связь некоторых природных полиморфизмов с общим понижением репарационной активности и, таким образом, могут выступать основой для создания маркерной панели полиморфизмов, ассоциированных с накоплением мутаций в ДНК.
Финансирование: Исследование проведено в рамках государственного задания № 121031300041-4, мутагенез, очистка и анализ свойств полиморфных вариантов POLβ выполнены в рамках проекта РНФ № 21-74-10103.

Natural polymorphisms of DNA base excision repair enzymes as a factor in the efficiency of damage removal

Alekseeva I.V., Kladova O.A., Kuznetsov N.A.*
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** nikita.kuznetsov@niboch.nsc.ru*

Key words: DNA repair; single-nucleotide polymorphism; enzymatic activity

Motivation and Aim: The stability of genetic information in a living cell is the key to its normal functioning. Today, there is no doubt that DNA damage plays a key role in the occurrence of various diseases, including cancer. The process of base excision repair (BER) involves the action of many enzymes and is responsible for removing a huge number of lesions from DNA. It should be noted that a decrease in the overall level of repair activity in humans can be caused by natural single nucleotide polymorphisms (SNPs) in the genes of DNA repair enzymes. Indeed, SNPs are widely distributed in the human population, and the relationship between SNPs and the etiology of several human

diseases remains an open area of research. At the same time, one of the pressing issues is to clarify the role of SNPs in the predisposition to the development of cancer and neurodegenerative diseases. It should also be noted that SNP-associated substitutions of amino acid residues of repair enzymes can lead to both the loss of functional properties of individual enzymes and disruption of protein-protein coordination of participants in the repair process. The purpose of this work was a systematic analysis of natural polymorphic variants of a number of human DNA base excision repair enzymes, namely DNA glycosylases SMUG1 and MBD4, AP endonuclease APE1 and DNA polymerase β.

Methods and Algorithms: Using the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/), it was found that thousands of SNPs were registered in the genes of DNA glycosylases SMUG1 and MBD4, AP endonuclease APE1 and DNA polymerase β. Bioinformation analysis methods have shown that some of them can affect the efficiency of enzymes. Molecular modeling of a number of selected mutant variants of enzymes was carried out to assess the possible effect of substitutions of amino acid residues on catalytic activity and the ability to form protein-protein complexes. To verify the results obtained, an experimental analysis of the enzymatic properties of a number of recombinant variants was carried out, which revealed a significant influence of selected natural substitutions both on the catalytic activity of enzymes and on their coordination with other participants in the repair process.

Results: Currently, there are a large number of bioinformatics approaches that make it possible to predict the effect of single nucleotide (SNP) substitutions on protein function. We analyzed known polymorphisms of DNA glycosylases SMUG1 and MBD4, AP endonuclease APE1 and human DNA polymerase β in order to identify polymorphic variants that can have a damaging effect on the functioning of these enzymes. Information on known single nucleotide polymorphisms was obtained from the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/snp). Particular attention was paid to polymorphisms that (1) led to a change in the class of amino acid residue, (2) were located in functionally important regions of the protein globule, and (3) were found in tumors. These criteria made it possible to select a number of SNP-associated substitutions in these enzymes for further analysis.

Analysis of the influence of natural polymorphisms on the enzymatic activity of APE1 (R221C, N222H, R237A, G241R, M270T, R274Q, P311S) and DNA glycosylases SMUG1 (G90C, P240H, N244S and N248Y) and MBD4 (S470L, G507S, R512W and H557D) showed that, if for some variants of DNA glycosylases an almost complete loss of enzymatic activity was detected (SMUG1 G90C and P240H, MBD4cat S470L, G507S and R512W), then in the case of APE1, apparently, the enzyme activity cannot be below the threshold value required to maintain the minimum level of repair activity sufficient for the survival of the organism. This difference in results may be due to the fact that in the case of DNA glycosylases, their substrate specificity can be duplicated by other enzymes of this class, while a decrease in APE1 activity leads to extremely negative consequences for the body.

When determining the activity of mutant POLβ variants in the transferase reaction, it was found that all the studied mutant forms (L19P, G66R, G118V, R149I, E154A, G189D, M236T, R254I, G274R, G290C, R333W) had a slower ability to fill 1 nt gaps in DNA, than the wild type enzyme. It should be noted that for mutant forms carrying the G274R and R333W substitutions, it was not possible to register the accumulation of

the reaction product, which indicates a significant degree of enzyme inactivation when substitutions of amino acid residues occur at positions G274 and R333.

The efficiency of coordinated interactions between APE1 and enzymes that perform their functions both upstream (DNA glycosylases: OGG1, UNG2, AAG) and downstream of APE1 (POLβ) in the BER process, as well as with protein factors (XRCC1 and PCNA) was determined by the ability effector proteins influence the activity of wild-type APE1 and a number of its natural polymorphic variants. It has been shown that amino acid residues Arg221 and Arg237, located on the surface of the protein globule, are important for the regulation of APE1 activity by other proteins of the repair complex, but at the same time have different effects on APE1 activity. While Asn222 apparently does not actively participate in contacts with all proteins tested, for example, the effector proteins OGG1, UNG2, and AAG lead to virtually no changes in the activity of this SNP variant. Introduction of amino acid substitutions into the DNA-binding center of APE1 leads to disruption of coordinated interactions with other BER participants. This conclusion is supported by the fact that both APE1 SNP variants M270T and R274Q are least sensitive to stimulation by the tested effector proteins compared to other APE1 polymorphic variants. The appearance of an additional positive charge in the protein globule in the case of G241R substitution leads to a decrease in the stimulation effect. AAG and PCNA have the greatest stimulating effect on the P311S variant. In addition, for the wild type uracil DNA glycosylases SMUG1 and MBD4cat, as well as their polymorphic variants SMUG1 N248Y and MBD4cat H557D, the stimulating effect of the catalytically inactive form of the AP endonuclease APE1 D210N was shown.

Determination of the influence of effector proteins (APE1, AAG, MBD4, NEIL1, NTHL1, OGG1 and TDG) on the accumulation of the product of the POLβ transferase reaction revealed that the addition of the APE1 enzyme to the reaction either completely inhibited or slowed down (WT POLβ, G66R, R149I) the accumulation of the product for all POLβ variants studied. In addition, for POLβ variants with a low polymerization constant (for example, L19P, G118V, G189D), the formation of products of APE1 exonuclease activity was recorded. DNA glycosylases AAG, NEIL1 and NTHL1 reduced the activity of the studied POLβ variants by more than 20 %. The presence of the MBD4 cat and TDG enzymes in the reaction had virtually no effect on the accumulation of the reaction product of individual POLβ variants, while the addition of the OGG1 enzyme led to a slight decrease in the accumulation of the reaction product, but did not inhibit the activity of the POLβ variants compared to the APE1 enzyme.

Conclusions: During the implementation of the project, an experimental analysis of 26 natural variants of SMUG1, MBD4, APE1 and POLβ was carried out, which revealed a significant effect of a number of natural substitutions both on the catalytic activity of the enzymes and on their coordination with other participants in the repair process. The data obtained confirm the connection of some natural polymorphisms with a general decrease in repair activity and, thus, can serve as the basis for creating a marker panel of polymorphisms associated with the accumulation of mutations in DNA.

Funding: The study was supported by the Russian Ministry of Science and Higher Education project No. 121031300041-4; the part of the work involving mutagenesis, enzyme purification, and kinetic analysis of POLβ was specifically funded by the Russian Science Foundation, grant No. 21-74-10103.

doi 10.18699/bgrs2024-11.1-02

Кинетические особенности взаимодействия APE1 и Polβ с ДНК и другими белками-участниками процесса эксцизионной репарации оснований ДНК

Бакман А.*, Кузнецов Н.

Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** art-bakman@yandex.ru*

Ключевые слова: эксцизионная репарация оснований; апуриновая/апиримидиновая эндонуклеаза 1; ДНК-полимераза β; белок-белковые взаимодействия; ферментативная кинетика

Мотивация и цель: Эксцизионная репарация оснований (BER) отвечает в живой клетке за репарацию таких повреждений ДНК, как апуриновые/апиримидиновые сайты (AP-сайты), модифицированные основания, одноцепочечные разрывы. BER – это сложный многостадийный процесс, в котором участвует ряд ферментов репарации и вспомогательных белковых факторов. В число ключевых ферментов BER у человека входят апуриновая/апиримидиновая эндонуклеаза 1 (APE1) и ДНК-полимераза β (Polβ). Несмотря на то что механизм реакции, катализируемой APE1, детально изучался во многих работах [1, 2], до сих пор остается дискуссионной роль некоторых аминокислотных остатков фермента в катализе, а также факторы, обеспечивающие субстратную специфичность APE1. Для полного понимания процесса BER следует выяснить не только механизмы взаимодействия ферментов репарации с ДНК, но и их координацию между собой, которая требуется для эффективного устранения повреждений ДНК. Координация BER осуществляется за счет множественных белок-белковых взаимодействий между участниками процесса и активно изучается исследователями в течение последних 20 лет [3, 4]. Тем не менее не все взаимодействия между участниками BER изучены на данный момент, а большинство белок-белковых взаимодействий не охарактеризованы в структурном и кинетическом аспектах.

Цель данной работы состояла в детальном кинетическом исследовании процесса согласованного взаимодействия апуриновой/апиримидиновой эндонуклеазы человека APE1 и ДНК-полимеразы β с ДНК в процессе эксцизионной репарации оснований.

В ходе исследований выполнялось определение функциональной роли аминокислотных остатков His309, Tyr171, Arg177 и Arg181 APE1 в механизме AP-эндонуклеазной реакции и уточнение молекулярно-кинетического механизма взаимодействия APE1 с модельными ДНК-субстратами; создание чувствительной системы регистрации процессов взаимодействия фермента с ДНК и другими белками BER на основе флуоресцентно меченной APE1; исследование взаимного влияния между APE1 и Polβ в процессе BER и анализ эффекта ДНК-гликозилаз на каталитическую активность Polβ.

Методы и алгоритмы: Для внесения мутаций в ген APE1 использовался метод сайт-направленного мутагенеза, для выделения и очистки белков использовались ионообменная и аффинная хроматографии, для анализа накопления продуктов

ферментативных реакций использовался метод электрофореза в денатурирующем полиакриламидном геле. Предстационарную кинетику взаимодействия ферментов с ДНК изучали методом «остановленной струи» с регистрацией флуоресценции остатков триптофана фермента, флуоресцеина в составе ДНК, 2-аминопурина в составе ДНК или красителя Alexa Fluor 488 в составе меченого белка. Обработка кинетических данных проводилась с помощью программ Origin 8.1 и DynaFit.

Результаты: Была изучена активность мутантных форм APE1 с заменами Y171F, H309A, R177A, R181A по отношению к субстрату, содержащему F-сайт (синтетический химически стабильный аналог AP-сайта). Показано, что замены Y171F и H309A не приводят к изменению профиля зависимости скорости катализа от pH. Исследование кинетики взаимодействия мутантных форм APE1 с ДНК-субстратом позволило установить влияние замен на стадии связывания фермента с ДНК, непосредственно расщепления фосфодиэфирной связи и диссоциации продукта. Показано, что замены Y171F и H309A снижают не только скорость расщепления ДНК на 3–4 порядка, но и скорость и эффективность связывания фермента с ДНК. Замена R177A не влияет на скорость стадии расщепления ДНК, но существенно облегчает диссоциацию комплекса фермента с продуктом. Замена R181A тоже облегчает диссоциацию продукта, но при этом еще и существенно снижает скорость стадии гидролиза фосфодиэфирной связи.

Использование метода химической модификации APE1 флуоресцентным красителем Alexa Fluor 488 в сочетании с сайт-направленным мутагенезом остатков цистеина в белке позволило получить белок, меченный по остатку Cys99, что было подтверждено масс-спектрометрическим анализом. Использование флуоресцентно меченной APE1 позволило зарегистрировать специфические процессы взаимодействия фермента с ДНК-субстратами в ходе эндонуклеазной реакции по изменению флуоресценции красителя в составе белка. Фазы изменения интенсивности флуоресценции Alexa Fluor 488 были соотнесены со стадиями процесса фермент-субстратного взаимодействия. Было изучено взаимодействие меченой APE1 с модельными 17- и 28-звенными ДНК-субстратами. Полученные данные существенно дополнили предложенную ранее модель фермент-субстратного взаимодействия путем детализации механизма диссоциации комплекса фермента с продуктом реакции.

Флуоресцентно меченная APE1 и 28-звенный ДНК-дуплекс были использованы в работе для того, чтобы изучить механизм влияния Polβ на активность APE1. Показано, что Polβ способствует связыванию APE1 с субстратом эндонуклеазной реакции и диссоциации ее из комплекса с продуктом (расщепленной ДНК). Было изучено влияние APE1 на кинетику катализируемого Polβ встраивания нуклеотида в ДНК-субстраты, моделирующие интермедиаты BER. В случае дуплекса с однонуклеотидной брешью добавление APE1 в реакционную смесь приводит к небольшому снижению скорости как связывания Polβ с ДНК, так и каталитической стадии. В случае субстрата, содержащего однонуклеотидную брешь с F-сайтом, концентрация APE1 не влияла на скорости связывания Polβ с ДНК и встраивания нуклеотида. Однако при высоких концентрациях APE1 на флуоресцентных кривых появлялась дополнительная фаза, предположительно, соответствующая связыванию APE1 с продуктом полимеразной реакции. Дополнительные эксперименты показали, что повышение концентрации Polβ приводит к ингибированию данной стадии, что свидетельствует о конкуренции

между двумя ферментами за связывание ДНК. Однако EMSA показал, что APE1 и Polβ могут связываться с ДНК-продуктом полимеразной реакции одновременно.
С целью изучения влияния ДНК-гликозилаз на нуклеотидилтрансферазную активность Polβ были выбраны 7 ДНК-гликозилаз, относящихся к разным структурным семействам и обладающих различными субстратными специфичностями: AAG, OGG1, NEIL1, NTHL1, MBD4, UNG2 и SMUG1. Катализируемое Polβ встраивание нуклеотида в ДНК-субстраты, имитирующие различные интермедиаты BER, в присутствии каждой из этих ДНК-гликозилаз было изучено в условиях кинетики одного оборота. Результаты показали, что ДНК-гликозилазы AAG, OGG1, NTHL1, MBD4, UNG2 и SMUG1 стимулируют активность Polβ по отношению к различным интермедиатам BER.
Выводы: Результаты исследования кинетики взаимодействия мутантных форм APE1 с ДНК-субстратами позволили уточнить функциональную роль остатков H309, Y171, R177 и R181 в механизме AP-эндонуклеазной реакции. Остатки Y171 и H309 не только критически важны для катализа эндонуклеазной реакции, но также участвуют в связывании APE1 с ДНК, что подтверждает их роль в стабилизации переходного состояния реакции за счет координации гидролизуемой фосфатной группы ДНК. R177 важен для связывания фермента с расщепленной ДНК, но не участвует в гидролизе фосфодиэфирной связи. R181 важен для стабилизации комплекса APE1 с расщепленной ДНК, но помимо этого способствует протеканию реакции гидролиза фосфодиэфирной связи.
С помощью созданной чувствительной модельной системы на основе флуоресцентно меченной APE1 зарегистрированы специфические процессы взаимодействия фермента с ДНК-субстратами различной длины, содержащими F-сайт, а также ДНК-опосредованные взаимодействия между APE1 и другими ферментами BER. Полученные данные позволили установить последовательность стадий диссоциации комплекса фермента с расщепленной ДНК. Показано, что в случае 17-звенных дуплексов диссоциация комплекса фермента с продуктом реакции инициируется высвобождением 5'-концевого олигонуклеотидного фрагмента, тогда как в случае более длинных модельных субстратов происходит согласованная диссоциация фермента и дуплексной ДНК, содержащей разрыв.
Изучены механизм влияния Polβ на взаимодействие APE1 с ДНК и механизм влияния APE1 на катализируемое Polβ встраивание нуклеотида в ДНК с однонуклеотидной брешью. Полученные результаты свидетельствуют о сложном механизме координации между этими двумя ферментами BER, выходящем за пределы модели «передачи эстафетной палочки».
Полученные в работе данные о влиянии различных ДНК-гликозилаз на нуклеотидилтрансферазную реакцию Polβ по отношению к различным интермедиатам BER свидетельствуют в пользу предположения, что ДНК-гликозилазы формируют контакты с Polβ в ходе осуществления полимеразной реакции в BER.
Финансирование: Исследование проведено в рамках государственного задания № 121031300041-4.

Kinetic features of protein-DNA and protein-protein interactions of APE1 and Polβ during the base excision repair

Bakman A.*, Kuznetsov N.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** art-bakman@yandex.ru*

Key words: base excision repair; apurinic/apyrimidinic endonuclease 1; DNA polumerase β; protein-protein interactions; enzyme kinetics

Motivation and Aim: In living cells Base excision repair (BER) is responsible for the repair of DNA damage such as apurinic/apyrimidinic sites (AP sites), modified bases, and single-strand breaks. BER is a complex, multistep process that involves a number of repair enzymes and auxiliary protein factors. Key BER enzymes in humans include apurinic/apyrimidinic endonuclease 1 (APE1) and DNA polymerase β (Polβ). The mechanism of the reaction catalyzed by APE1 has been studied in detail in many studies [1, 2]. However, the role of some amino acid residues of the enzyme in catalysis and the factors ensuring the substrate specificity of APE1 still remain controversial. To fully understand the BER process, it is necessary to elucidate not only the mechanisms of interaction of repair enzymes with DNA, but also their coordination with each other, which is required for effective repair of DNA damage. BER is coordinated by multiple protein-protein interactions between participants in the process and has been actively studied by researchers over the past 20 years [3, 4]. However, not all interactions between BER participants have been studied to date, and most protein-protein interactions have not been characterized in structural and kinetic aspects.
The aim of this work was to perform a detailed kinetic study of the coordinated interaction of human apurinic/apyrimidinic endonuclease APE1 and DNA polymerase β with DNA during the base excision repair.
In the course of the studies we determined the functional role of amino acid residues His309, Tyr171, Arg177 and Arg181 of APE1 in the mechanism of AP-endonuclease reaction and clarified the molecular kinetic mechanism of interaction of APE1 with model DNA substrates. Also we created a sensitive system for recording the interaction of the enzyme with DNA and other BER proteins based on fluorescently labeled APE1. Finally, we studied the mutual influence between APE1 and Polβ during the BER process and analyzed the effect of DNA glycosylases on the catalytic activity of Polβ.
Methods and Algorithms: Site-directed mutagenesis was used to introduce mutations into the APE1 gene; ion-exchange and affinity chromatography were used to isolate and purify proteins; electrophoresis in a denaturing polyacrylamide gel was used to analyze the accumulation of enzymatic reaction products, The pre-steady-state kinetics of enzyme-DNA interaction was studied by the “stopped flow” technique with registrating the fluorescence of tryptophan residues of the enzyme, fluorescein in DNA, 2-amino-purine in DNA or Alexa Fluor 488 dye in labeled protein. Kinetic data were processed using Origin 8.1 and DynaFit programs.
Results: The activities of the APE1 mutant forms with substitutions Y171F, H309A, R177A, and R181A toward substrates containing F-site (synthetic chemically stable analog of AP-site) were studied. It is shown that substitutions Y171F and H309A do not lead to a change in the profile of the pH dependence of the catalysis rate. The study of

the kinetics of interaction of mutant forms of APE1 with DNA-substrate allowed us to establish the influence of substitutions on the stages of enzyme binding to DNA, cleavage of the phosphodiester bond and dissociation of the product. It was shown that substitutions Y171F and H309A not only reduce the rate of DNA cleavage by 3–4 orders of magnitude, but also reduce the rate and efficiency of enzyme-DNA binding. Substitution R177A does not affect the rate of the DNA cleavage step, but significantly facilitates the dissociation of the enzyme-product complex. Substitution R181A also facilitates dissociation of the product, but also significantly reduces the rate of the hydrolysis stage of the phosphodiester bond.

The chemical modification of APE1 by the fluorescent dye Alexa Fluor 488 in combination with site-directed mutagenesis of cysteine residues in the protein allowed us to obtain a protein labeled by the Cys99 residue, which was confirmed by mass spectrometric analysis. The use of fluorescently labeled APE1 allowed us to register specific processes of interaction of the enzyme with DNA-substrates during the endonuclease reaction by changes in the fluorescence of the dye in the protein composition. The phases of the fluorescence intensity change of Alexa Fluor 488 were related to the stages of the enzyme-substrate interaction process. The interaction of labeled APE1 with model 17- and 28-nt DNA substrates was studied. significantly supplemented the previously proposed model of the enzyme-substrate interaction by detailing the mechanism of dissociation of the enzyme complex with the product.

Fluorescently labeled APE1 and the 28-nt DNA duplex were used in this work to study the mechanism of Polβ's influence on APE1 activity. Polβ was shown to facilitate the binding of APE1 to the substrate of the endonuclease reaction and its dissociation from the complex with the product (cleaved DNA). The effect of APE1 on the kinetics of Polβ-catalyzed nucleotide incorporation into DNA substrates modeling BER intermediates was studied. In the case of a duplex with a single-nucleotide gap, the addition of APE1 to the reaction mixture results in a slight decrease in the rate of both Polβ binding to DNA and the catalytic step. In the case of a substrate containing a single-nucleotide gap with an F-site, the concentration of APE1 did not affect the rates of Polβ binding to DNA and nucleotide incorporation. However, at high APE1 concentrations, an additional phase appeared in the fluorescence curves, presumably corresponding to the binding of APE1 to the product of the polymerase reaction. Additional experiments showed that increasing the concentration of Polβ resulted in inhibition of this phase, indicating competition between the two enzymes for DNA binding. However, EMSA has shown that APE1 and Polβ bind to the DNA product of the polymerase reaction simultaneously.

In order to study the effect of DNA glycosylases on the nucleotidyltransferase activity of Polβ, 7 DNA glycosylases belonging to different structural families and with different substrate specificities were selected: AAG, OGG1, NEIL1, NTHL1, MBD4, UNG2, and SMUG1. Polβ-catalyzed nucleotide incorporation into DNA substrates mimicking different BER intermediates in the presence of each of these DNA glycosylases was studied under single turnover kinetics conditions. The results have shown that DNA glycosylases AAG, OGG1, NTHL1, MBD4, UNG2, and SMUG1 stimulate Polβ activity toward different BER intermediates.

Conclusion: The results of studying the kinetics of interaction of the APE1 mutant forms APE1 with DNA substrates allowed us to clarify the functional role of residues H309, Y171, R177, and R181 in the mechanism of the AP-endonuclease reaction. Residues Y171 and H309 are not only critical for catalysis of the endonuclease reaction, but also

participate in the binding of APE1 to DNA, which confirms their role in stabilizing the reaction transition state by coordinating the hydrolyzable phosphate group of DNA. R177 is important for the binding of the enzyme to cleaved DNA but is not involved in the hydrolysis of the phosphodiester bond. R181 is important for stabilization of the APE1 complex with cleaved DNA, but also contributes to the reaction of phosphodiester bond hydrolysis.

Using the created by us sensitive model system based on fluorescently labeled APE1, specific processes of interaction of the enzyme with DNA substrates of different length containing F-site, as well as DNA-mediated interactions between APE1 and other BER enzymes were registered. The data obtained allowed to establish the sequence of stages of dissociation of the enzyme complex with cleaved DNA. It was shown that in the case of our model 17-nt duplexes, dissociation of the enzyme complex with the reaction product is initiated by the release of the 5'-terminal oligonucleotide fragment, while in the case of longer model substrates there is a coordinated dissociation of the enzyme and the duplex DNA containing the break.

The mechanism of Polβ's influence on the interaction of APE1 with DNA Polβ and the mechanism of APE1's influence on the Polβ-catalyzed nucleotide insertion into DNA with a single-nucleotide gap were studied. These results indicate a complex mechanism of coordination between these two BER enzymes that goes beyond the “baton passing” model.

The data obtained in this work regarding the effect of different DNA-glycosylases on the Polβ's nucleotidyltransferase activity toward different BER intermediates support the assumption that DNA-glycosylases form contacts with Polβ during the polymerase reaction in BER.

Funding: The research was conducted under the state assignment No. 121031300041-4.

Список литературы/References

1. Mol C.D., Izumi T., Mitra S., Tainer J.A. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature*. 2000;403:451-456
2. Freudenthal B.D., Beard W.A., Cuneo M.J., Dyrkheeva N.S., Wilson S.H. Capturing snapshots of APE1 processing DNA damage. *Nat Struct Mol Biol*. 2015;22:924-931
3. Moor N.A., Lavrik O.I. Protein – Protein Interactions in DNA Base Excision Repair. *Biochemistry (Moscow)*. 2018;83:411-422
4. Dianov G.L., Hübscher U. Mammalian base excision repair: The forgotten archangel. *Nucleic Acids Res*. 2013;41:3483-3490

doi 10.18699/bgrs2024-11.1-03

Взаимодействие РНКазы H1 из хлоропластов *A. thaliana* (AtRNH1C) с модельными R-петлями различной структуры

Гаврилова А.*, Косарев Ю., Микушина Е., Кузнецов Н., Кузнецова А.
Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** anaroslakova@gmail.com*

Ключевые слова: РНКаза H1; *A. thaliana*; конфликт транскрипции-репликации

Мотивация и цель: Транскрипция и репликация являются важнейшими молекулярно-биологическими процессами, постоянно протекающими в клетках. Оба процесса происходят на одной и той же матрице ДНК, в связи с чем вероятность встреч между этими важнейшими клеточными механизмами, ответственными за экспрессию генов и дупликацию генома, достаточно высока, что может вызвать конфликт транскрипции-репликации (Transcription-Replication Conflict, TRC) [1]. Следует отметить, что одним из структурных элементов TRC является формирование R-петель, структур, состоящих из гетеродуплекса ДНК:РНК и содержащих выпетленный одноцепочечный фрагмент ДНК.

Для того, чтобы установить роль отдельных белков, участвующих во взаимодействии с R-петлями, следует учитывать не только процесс транскрипции, в котором эти структуры формируются, но также необходимо отметить биологически важные процессы, на протекание которых эти структуры оказывают влияние [2–4]. Так, с одной стороны, R-петли играют важную роль в регуляции экспрессии генов, влияя на терминацию транскрипции, модификацию ДНК и топологию хроматина, а также участвуют в процессах репликации ДНК, репарации ДНК и рекомбинации с переключением классов иммуноглобулинов. С другой стороны, R-петли могут приводить к повреждению ДНК, поскольку одноцепочечные участки ДНК, образующиеся в результате гибридизации одной из цепей с РНК, подвержены повреждениям, что, в свою очередь, может приводить как к мутагенезу, так и к разрывам цепи, и, таким образом, вызывает нестабильность генома.

Считается, что R-петли являются короткоживущими структурами, однако в некоторых условиях они могут быть значительно стабилизированы [5]. Поскольку R-петли могут приводить к повреждению ДНК и нестабильности генома, уровень R-петель в клетках строго контролируется [6]. Хотя, как прокариотические, так и эукариотические клетки имеют множество механизмов координации процессов транскрипции и репликации, в настоящее время очевидно, что конфликты могут возникать из-за аберрантной регуляции транскрипции и преждевременной пролиферации [7, 8].

Актуальными вопросами в настоящее время остаются как понимание клеточных последствий TRC, так и идентификация клеточных регуляторных механизмов, ответственных за деблокирование репликации в процессе TRC, а также изучение молекулярных механизмов, которые используются клетками для предотвращения негативных эффектов R-петель, формирующихся в TRC. Один из механизмов разрешения TRC в геномной ДНК осуществляется с помощью членов семейства ферментов, обладающих активностью РНКазы H. Показано, что для поддержания

стабильности генома хлоропластов *A. thaliana* и уменьшения негативного эффекта R-петель требуется согласованное действие РНКазы H (AtRNH1C), ДНК-гиразы (AtGyr) и праймазы (ATH) [9–10]. В связи с этим в рамках настоящей работы с целью реконструкции *in vitro* молекулярных взаимодействий между белками-участниками процесса TRC проведен кинетический анализ особенностей взаимодействия РНКазы H1 AtRNH1C из хлоропластов *Arabidopsis thaliana* с модельными R-петлями различной структуры.

Методы и алгоритмы: Созданы модельные синтетические R-петли, содержащие флуоресцентные метки в различных положениях ДНК- и РНК-цепей, для детекции белково-нуклеиновых взаимодействий между отдельными ферментами и/или их комплексами с белками-регуляторами с модельными R-петлями. Методы флуоресцентного титрования и микроскопического термофореза были использованы для определения констант связывания и термодинамических характеристик белково-нуклеиновых взаимодействий между отдельными ферментами, а также комплексами ферментов с другими участниками ферментативного процесса и белками-регуляторами в составе модельных R-петель.

Результаты: Анализ конформационной подвижности фермента и субстрата при их взаимодействии в случае изучения взаимодействий между AtRNH1C и модельными R-петлями представляет собой нетривиальную задачу, требующую дизайна гибридного гетеродуплекса ДНК:РНК, содержащего выпетленный одноцепочечный фрагмент ДНК и флуоресцентный маркер конформационных изменений. На первом этапе работы нами создан набор ДНК-дуплексов, содержащих в центральной части некомплементарный участок (пузырь) длиной 10, 11, 12, 16, 21 и 31 нуклеотидов. При этом с транскрибируемой цепью формировали гетеродуплекс с помощью рибоолигонуклеотида длинной от 10 до 37 нуклеотидов, таким образом, чтобы 3'-концевой нуклеотид РНК-праймера располагался в разных положениях относительно некомплементарного участка ДНК-дуплекса. На 3'-конце РНК-праймера был расположен либо аденозин, либо его флуоресцентный аналог – 2-аминопурин (aPu). Кроме того, был синтезирован ряд матричных и нематричных цепей ДНК-дуплекса, содержащих остаток aPu на разном расстоянии от некомплементарного участка. Остаток aPu был выбран в качестве флуоресцентного маркера конформационных превращений в R-петле на основании его чувствительности к окружению. Так, по сравнению со свободным нуклеотидом интенсивность флуоресценции aPu в составе олигонуклеотида снижается из-за стэкинга с соседними основаниями, кроме того, образование дуплексной структуры из одноцепочечных олигонуклеотидов, как правило, приводит к дополнительному уменьшению интенсивности флуоресценции aPu за счет стабилизации стэкинг-взаимодействий с соседними азотистыми основаниями. Несмотря на то, что aPu является аналогом аденозина и способен формировать две водородные связи с тимином в комплементарной цепи, не приводя к структурным изменениям дуплекса, для всех разработанных РНК-праймеров, содержащих 3'-концевой остаток aPu, в качестве контроля использовали олигонуклеотид, содержащий 3'-концевой аденозин.

Для формирования R-петель РНК-праймеры отжигались с матричной цепью ДНК в буферном растворе 40 мМ Tris-HCl pH 7.9, 40 мМ KCl и инкубировались с нематричной цепью ДНК 35 мин при 37 °C либо 30 мин при 25 °C в том же буферном растворе. Конечная смесь содержала 1 мкМ РНК-праймер, 1.1 мкМ

матричную цепь ДНК, 5 мкМ нематричную цепь ДНК. Была проверена стабильность модельных R-петель в экспериментальных условиях, необходимых для исследования ферментативной активности РНКазы H.
Для анализа процесса деградации РНК-праймера в полиакриламидном геле использовали рибоолигонуклеотиды 5'-FAM-CUCACAUCCAGAGAGCX (X – аденозин либо aPu), содержащие на 5'-конце краситель флуоресцеин (FAM), позволяющий визуализировать продукты деградации РНК-праймера ферментом AtRNH1C.
Все разработанные модельные R-петли использовали для определения кинетических параметров взаимодействия РНКазы H1 AtRNH1C из хлоропластов *Arabidopsis thaliana*. Для каждого субстрата получена серия кинетических кривых, в которой концентрация субстратов была постоянной, а концентрацию фермента варьировали, что позволило оценить эффективность деградации РНК-праймера. Поскольку общая эффективность расщепления РНК-праймера зависит как от скорости формирования фермент-субстратного комплекса, так и от скорости каталитической стадии в рамках работы определены константы связывания AtRNH1C с модельными R-петлями методом микроскопического термофореза.
Выводы: В работе проведен кинетический анализ процесса взаимодействия РНКазы H1 AtRNH1C из хлоропластов *Arabidopsis thaliana* с модельными R-петлями различной структуры. Полученные данные выявили особенности формирования фермент-субстратного комплекса и каталитической стадии в зависимости от структуры модельных R-петель.
Финансирование: Исследование поддержано грантом РНФ № 23-44-00064.

Interaction of chloroplast RNase H1 from *A. thaliana* (AtRNH1C) with model R-loops of different structure

Gavrilova A.*, Kosarev Y., Mikushina E., Kuznetsov N., Kuznetsova A.
Institute of chemical biology and fundamental medicine, SB RAS, Novosibirsk, Russia
** anaroslakova@gmail.com*

Key words: RNase H1; Arabidopsis thaliana; R-loops; transcription-replication conflict

Motivation and Aim: Transcription and replication are the most important molecular biological processes that occur at high frequency in cells. As they share the same DNA template, a high incidence of encounters is expected between these most important cellular machineries responsible for gene expression and genome duplication, which can cause a transcription-replication conflict (TRC) [1]. It should be noted that one of the structural elements of TRC is the formation of R-loops, structures consisting of a DNA:RNA heteroduplex and containing a looped single-stranded DNA fragment. It is necessary to take into account the role of individual proteins involved in interaction with R-loops as well as influence of which these structures at the course of the biologically important processes [2–4].
Thus, on the one hand, R-loops play an important role in the regulation of gene expression, influencing transcription termination, DNA modification and chromatin topology, and also participate in the processes of DNA replication, DNA repair and recombination with immunoglobulin class switching. On the other hand, R-loops can lead to DNA damage, since single-stranded DNA regions formed as a result of

hybridization of one of the strands with RNA are susceptible to damage, which, in turn, can lead to both mutagenesis and strand breaks, and thus causing genomic instability.
It is believed that R-loops are short-lived structures, but under certain conditions they can be significantly stabilized [5]. Since R-loops can lead to DNA damage and genome instability, the level of R-loops in cells is strictly controlled [6]. Although both prokaryotic and eukaryotic cells have multiple mechanisms to coordinate transcription and replication processes, it is now clear that conflicts can arise due to aberrant transcriptional regulation and premature proliferation [7, 8].
Current issues remain both understanding the cellular consequences of TRC and identifying the cellular regulatory mechanisms responsible for unblocking replication in the TRC process, as well as studying the molecular mechanisms that cells use to prevent the negative effects of R-loops formed in TRC. One of the mechanisms for resolving TRC in genomic DNA is carried out with the help of members of a family of enzymes with RNase H activity. It has been shown that to maintain the stability of the genome of *A. thaliana* chloroplasts and reduce the negative effect of R-loops, the coordinated action of RNase H1 (AtRNH1C), DNA gyrase (AtGyr) and primase (ATH) is required [9–10]. In this regard, within the framework of this work, in order to reconstruct *in vitro* molecular interactions between proteins participating in the TRC process, a kinetic analysis of the interaction features of RNase H1 AtRNH1C from *Arabidopsis thaliana* chloroplasts with model R-loops of various structures was carried out.
Methods and Algorithms: Model synthetic R-loops containing fluorescent tags in various positions of DNA and RNA chains have been created to detect protein-nucleic acid interactions between individual enzymes and/or their complexes with regulatory proteins with model R-loops. Fluorescent titration and microscopic thermophoresis methods were used to determine binding constants and thermodynamic characteristics of protein-nucleic acid interactions between individual enzymes, as well as complexes of enzymes with other participants in the enzymatic process and regulatory proteins in model R-loops.
Results: Analysis of the conformational mobility of the enzyme and substrate during their interaction in the case of studying interactions between AtRNH1C and model R-loops is a non-trivial task that requires the design of a hybrid DNA:RNA heteroduplex containing a looped single-stranded DNA fragment and a fluorescent marker of conformational changes. At the first stage of our work, we created a set of DNA duplexes containing in the central part a non-complementary region (bubble) with a length of 10, 11, 12, 16, 21 and 31 nucleotides. In this case, a heteroduplex was formed with the transcribed strand using a ribooligonucleotide from 10 to 37 nucleotides in length, so that the 3'-terminal nucleotide of the RNA primer was located in different positions relative to the non-complementary region of the DNA duplex. At the 3' end of the RNA primer, either adenosine or its fluorescent analogue, 2-aminopurine (aPu), was located. In addition, a number of template and non-template strands of the DNA duplex were synthesized, containing an aPu residue at different distances from the non-complementary region. The aPu residue was chosen as a fluorescent marker of conformational transformations in the R-loop based on its sensitivity to the environment. Thus, compared to a free nucleotide, the fluorescence intensity of aPu in the oligonucleotide decreases due to stacking with neighboring bases; in addition, the formation of a duplex structure from single-stranded oligonucleotides, as a rule, leads to an additional decrease in the fluorescence intensity of aPu due to the stabilization of stacking interactions with neighboring bases. Despite the fact that aPu is an analogue of

adenosine and is capable of forming two hydrogen bonds with thymine in the complementary strand without leading to structural changes in the duplex, for all designed RNA primers containing the 3'-terminal residue of aPu, an oligonucleotide containing 3'-terminal adenosine.

To form R-loops, RNA primers were annealed to the template DNA strand in a buffer solution of 40 mM Tris-HCl pH 7.9, 40 mM KCl and incubated with the non-template DNA strand for 35 min at 37 °C or 30 min at 25 °C, including the same buffer solution. The final mixture contained 1 μM RNA primer, 1.1 μM template DNA strand, and 5 μM non-template DNA strand. The stability of model R-loops was tested under experimental conditions necessary for studying the enzymatic activity of RNase H.

To analyze the process of RNA primer degradation in a polyacrylamide gel, we used ribooligonucleotides 5'-FAM-CUCACAUCCAGAGAGCX (X – adenosine or aPu), containing the fluorescein dye (FAM) at the 5' end, which allows visualization of the products of RNA primer degradation by the AtRNH1C enzyme.

All developed model R-loops were used to determine the kinetic parameters of the interaction of RNase H1 AtRNH1C from *Arabidopsis thaliana* chloroplasts. For each substrate, a series of kinetic curves was obtained in which the concentration of the substrates was constant, and the concentration of the enzyme was varied, which made it possible to evaluate the efficiency of degradation of the RNA primer. Since the overall efficiency of RNA primer cleavage depends both on the rate of formation of the enzyme-substrate complex and on the rate of the catalytic stage, the binding constants of AtRNH1C to model R-loops were determined using microscopic thermophoresis.

Conclusion: The work carried out a kinetic analysis of the interaction of RNase H1 AtRNH1C from *Arabidopsis thaliana* chloroplasts with model R-loops of various structures. The data obtained revealed the peculiarities of the formation of the enzyme-substrate complex and the catalytic stage depending on the structure of the model R-loops.

Funding: The study is supported by the Russian Science Foundation, grant No. 23-44-00064.

Список литературы/References

1. García-Muse T., Aguilera A. Transcription–replication conflicts: how they occur and how they are resolved. *Nat Rev Mol Cell Biol*. 2016;17:553-563. doi 10.1038/nrm.2016.88
2. Costantino L, Koshland D. The Yin and Yang of R-loop biology. *Curr Opin Cell Biol.* 2015;34:39-45. doi 10.1016/j.ceb.2015.04.008
3. Brambati A., Zardoni L., Nardini E., Pellicioli A., Liberi G. The dark side of RNA:DNA hybrids. *Mutat Res Rev Mutat Res*. 2020;784:108300. doi 10.1016/j.mrrev.2020.108300
4. Mackay R.P., Xu Q., Weinberger P.M. R-Loop Physiology and Pathology: A Brief Review. *DNA Cell Biol*. 2020;39:1914-1925. doi 10.1089/dna.2020.5906
5. Crossley M.P., Bocek M., Cimprich K.A. R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell.* 2019;73:398-411. doi 10.1016/j.molcel.2019.01.024
6. Belotserkovskii B.P., Tornaletti S., D'Souza A.D., Hanawalt P.C. R-loop generation during transcription: Formation, processing and cellular outcomes. *DNA Repair*. 2018;71:69-81. doi 10.1016/j.dnarep.2018.08.009
7. Goehring L., Huang T.T., Smith D.J. Transcription-replication conflicts as a source of genome instability. *Annu Rev Genet*. 2023;57:157-179. doi 10.1146/annurev-genet-080320-031523
8. Lalonde M., Trauner M., Werner M., Hamperl S. Consequences and Resolution of Transcription – Replication Conflicts. *Life (Basel)*. 2021;11(7):637
9. Zhao H., Zhu M., Limbo O., Russell P. RNase H eliminates R-loops that disrupt DNA replication but is nonessential for efficient DSB repair. *EMBO Rep*. 2018;19:e45335. doi 10.15252/embr.201745335
10. Zhang W., Yang Z., Wang W., Sun Q. Primase promotes the competition between transcription and replication on the same template strand resulting in DNA damage. *Nat Comm*. 2024;15:73. doi 10.1038/s41467-023-44443-0

doi 10.18699/bgrs2024-11.1-04

Политенные хромосомы как инструмент для поиска факторов, влияющих на эффективность инициации репликации ДНК

Колесникова Т.Д.[1]*, Воробьева Н.Е.[2], Черендина К.П.[3], Балантаева М.Н.[3], Довгань В.В.[4], Шуберт В.[5], Жимулев И.Ф.[1]

[1] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[2] *Институт биологии гена РАН, Москва, Россия*

[3] *Новосибирский государственный университет, Новосибирск, Россия*

[4] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[5] *Институт генетики растений и исследований сельскохозяйственных культур имени Лейбница, Гатерслебен, Германия*

* *kolesnikova@mcb.nsc.ru*

Ключевые слова: *Drosophila*; политенные хромосомы; инициация репликации; эффективность инициации репликации; EcR, 3D-SIM

Мотивация и цель: В эукариотических клетках в каждом клеточном цикле используется лишь небольшая доля потенциальных сайтов инициации репликации ДНК, при этом каждый сайт характеризуется определенной вероятностью (эффективностью) инициации репликации. Политенные хромосомы дрозофилы в сочетании с микроскопией сверхвысокого разрешения дают уникальную возможность для анализа вероятностного характера инициации репликации на ультраструктурном уровне. Ранее мы показали, что инициация репликации в политенных хромосомах происходит преимущественно в интервалах между наиболее компактными дисками политенных хромосом, а внутри этих интервалов инициация носит вероятностный характер [1, 2]. Каждый интервал действует как зона инициации репликации, где на каждой нити ДНК репликация инициируется лишь на одном из потенциальных ориджинов репликации. Эти зоны существенно различаются по вероятности инициации репликации в единицу времени. Целью работы было выявить зоны, инициирующие репликацию максимально эффективно, и выяснить, какие факторы влияют на эффективность инициации в политенных хромосомах слюнных желез *Drosophila melanogaster*.

Методы и алгоритмы: Мы анализировали сайты ранней эффективной инициации репликации у личинок дикого типа, а также при индукции эктопической S фазы слабым тепловым шоком у личинок линий дрозофил, несущих трансген *hsp70-CycE* [3]. Для визуализации репликации использовали импульсное включение 5-этинил-2'-дезоксиуридин (EdU). Для визуализации в политенных хромосомах активной транскрипции и связывания транскрипционных факторов использовали иммуноокрашивание политенных хромосом соответствующими антителами. Для выявления цитологических сигналов в политенных хромосомах с максимальным разрешением была использована микроскопия сверхвысокого разрешения со структурированным освещением (3D-structured illumination microscopy, 3D-SIM).

Результаты и выводы: Мы выявили около 80 сайтов очень эффективной инициации ранней репликации, в которых инициация происходит на большинстве нитей политенной хромосомы в первые несколько минут S фазы. Некоторые

хромосомные локусы демонстрировали эффективную инициацию репликации на всех анализируемых стадиях личиночного развития, тогда как другие выступали в роли эффективных сайтов инициации репликации только на определенных стадиях. Многие сайты соответствовали локусам генов, транскрипция которых регулируется гормоном экдизоном, и сайтам связывания антител к белку рецептора экдизона (EcR). При этом у личинок дикого типа инициация репликации и активная транскрипция в этих локусах происходят на разных этапах развития, что может быть связано с механизмом избегания конфликта между репликацией и транскрипцией в этих локусах. Исключением является группа генов *sgs*, производящих секрет слюнной железы, максимальная экспрессия которых совпадает со стадией развития, когда многие клетки слюнных желез вступают в S фазу. Локализация выявленных сайтов на геномной карте и анализ распределения факторов транскрипции и модификаций хроматина позволил сделать вывод, что в клетках слюнных желез наиболее эффективные ориджины/кластеры ориджинов появляются в участках хромосом, несущих энхансерные маркеры.

Финансирование: Исследование поддержано РНФ (№ 24-14-00133).

Polytene chromosomes as a tool to identify factors influencing the efficiency of DNA replication initiation

Kolesnikova T.D.[1]*, Vorobyeva N.E.[2], Cherendina K.P.[3], Balantaeva M.N.[3], Dovgan V.V.[4], Schubert V.[5], Zhimulev I.F.[1]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[2] *Institute of Gene Biology, RAS, Moscow, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

[4] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[5] *Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany*

* *kolesnikova@mcb.nsc.ru*

Key words: *Drosophila*; polytene chromosome; replication initiation; origin efficiency; EcR

Motivation and Aim: In eukaryotic cells, only a small fraction of potential DNA replication initiation sites is used in each cell cycle, and each site is characterized by a certain probability (efficiency) of replication initiation. *Drosophila* polytene chromosomes, combined with super-resolution microscopy, provide a unique opportunity to analyze the probabilistic nature of replication initiation at the ultrastructural level. We have previously shown that the initiation of replication in polytene chromosomes occurs predominantly in the intervals between the most compact polytene chromosome bands, and within these intervals the initiation is probabilistic [1, 2]. Each interval acts as a replication initiation zone, where on each DNA strand, replication is initiated at only one of the potential origins of replication. These zones differ significantly in the probability of replication initiation per time unit. The goal of the work was to identify the zones that initiate replication very efficiently and to find out what factors influence the efficiency of initiation in *D. melanogaster* salivary gland polytene chromosomes.

Methods and Algorithms: We analyzed very efficient early replication initiation sites in wild-type polytene chromosomes, and in chromosomes of hsp70-CycE larvae [3] after ectopic S phase induction [2]. To visualize replication, 5-ethynyl-2′-deoxyuridine (EdU) pulses were applied. To visualize active transcription and binding of transcription factors in polytene chromosomes, immunostaining with appropriate antibodies was performed. To analyze the ultrastructure of chromatin and immunosignals at the super-resolution level spatial structured illumination microscopy (3D-SIM) was applied.
Results and Conclusions: We identified ~80 very efficient early replication initiation sites, in which initiation occurs on most polytene chromosome strands during the first few minutes of the S phase. Some chromosomal loci demonstrated efficient initiation of replication at all stages of larval development analyzed, whereas others acted as effective replication initiation sites only at certain stages. Many of the sites corresponded to gene loci whose transcription is regulated by the hormone ecdysone and stained by antibodies to the ecdysone receptor (EcR) protein. Moreover, in wild-type larvae, the initiation of replication and active transcription in these loci occur at different developmental stages, which may be associated with a mechanism to avoid a conflict between replication and transcription. An exception is the group of *sgs* genes that produce salivary gland secretions. Here the maximum transcription coincides with the developmental stage when many salivary gland cells enter the S phase. The localization of the identified efficient early replication initiation sites on the genomic map and the analysis of the distribution of transcription factors and chromatin modifications led to the conclusion that in salivary gland cells, the most efficient replication origins occur in chromosomal regions containing transcriptional enhancers.
Funding: The study is supported by RSF (No. 24-14-00133).

Список литературы/References

1. Kolesnikova T.D., Goncharov F.P., Zhimulev I.F. Similarity in replication timing between polytene and diploid cells is associated with the organization of the Drosophila genome. *PLoS One*. 2018;13(4):e0195207
2. Kolesnikova T.D., Pokholkova G.V., Dovgan V.V., Zhimulev I.F., Schubert V. Super-resolution microscopy reveals stochastic initiation of replication in Drosophila polytene chromosomes. *Chromosome Res*. 2022;30(4):361-383
3. Knoblich J.A., Sauer K., Jones L., Richardson H., Saint R., Lehner C.F. Cyclin E controls S phase progression and its down-regulation during *Drosophila* embryogenesis is required for the arrest of cell proliferation. *Cell*. 1994;77(1):107-120

doi 10.18699/bgrs2024-11.1-05

Положение повреждения ДНК в составе нуклеосом влияет на HPF1-зависимую модификацию гистонов ферментами PARP1 и PARP2

Кургина Т.*, Моор Н., Кутузов М., Лаврик О.
Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** t.a.kurgina@gmail.com*

Ключевые слова: репарация ДНК; PAR-илирование; PARP1; PARP2; HPF1; гистоны

Мотивация и цель: Поли(ADP-рибозил)ирование (PAR-илирование) представляет собой посттрансляционную модификацию белков, которую осуществляют ферменты, известные как поли(ADP-рибоза)полимеразы (PARP). PAR-илирование играет ключевую роль в регуляции различных клеточных процессов, включая репарацию ДНК, организацию структуры хроматина, экспрессию генов, репликацию ДНК, процессинг РНК и биогенез рибосом [1]. Основные ферменты, катализирующие PAR-илирование в ядре и синтез протяженного полимера поли(ADP-рибозы) (PAR), – два ДНК-зависимых фермента семейства PARP: PARP1 и PARP2 [2]. Функции этих ферментов в регуляции репарации ДНК существенно пересекаются, оба они участвуют в эксцизионной репарации оснований (BER) и взаимодействуют с ключевыми компонентами этого процесса, такими как ДНК-полимераза-β, XRCC1 и APE1 [3]. Ферменты PARP1 и PARP2 взаимодействуют с молекулами ДНК, имитирующими интермедиаты коротко- и длиннозаплаточного путей BER, что было показано с использованием синтетических ДНК-дуплексов, содержащих повреждения, и нуклеосом [4]. Здесь мы показали различия во взаимодействии ферментов PARP1 и PARP2 с повреждениями ДНК – интермедиатами BER на модельных нуклеосомах, а также влияния положения повреждения на модификацию гистонов данными ферментами.
Методы и алгоритмы: Электрофорез в ПААГ по методу Леммли, флуоресцентная спектрофотомерия.
Результаты: Несмотря на обширные исследования аффинности PARP1 и PARP2 к различным формам ДНК повреждений, включая однонуклеотидные бреши в контексте нуклеосомных коровых частиц (nucleosome core particle, NCP), детального рассмотрения взаимодействия этих ферментов с порождениями ДНК, в зависимости от их удаленности от 5'-конца ДНК, не проводилось. В данном исследовании для оценки величин EC_{50}, отражающих аффинность PARP1 и PARP2 к различным активаторам ДНК, было использовано измерение анизотропии флуоресценции. Мы показали, что приближение бреши к 5'-концу ДНК приводило к увеличению значений EC_{50} для PARP1, но не для PARP2. Снижение аффинности может быть связано с конкуренцией молекул PARP1 за связывание с двумя близкорасположенными повреждениями: тупым концом ДНК и однонуклеотидной брешью. В то же время PARP2 специфически взаимодействует с брешью с высокой аффинностью, и данное повреждение является единственным сайтом специфического взаимодействия PARP2 с NCP.

Дополнительно было проанализировано количество и распределение PAR, синтезированного PARP1 и PARP2 в присутствии фактора PAR-илирования гистонов HPF1. Мы показали, что модификация гистонов зависит скорее от положения бреши относительно гистонового ядра, нежели от удаленности бреши от 5'-конца ДНК.
Выводы: Присутствие бреши в структуре ДНК в нуклеосоме влияет на HPF1-зависимое PAR-илирование гистонов. Сродство PARP1 к однонуклеотидной бреши в ДНК зависит от удаленности бреши от 5'-конца ДНК.
Финансирование: Исследование поддержано грантом РНФ № 22-74-10059.

DNA damage position in the nucleosomes affects the HPF1-dependent modification of histones by PARP1 and PARP2

Kurgina T.*, Moor N., Kutuzov M., Lavrik O.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** t.a.kurgina@gmail.com*

Key words: DNA repair; PARylation; PARP 1; PARP2; HPF1; histones

Motivation and Aim: Poly(ADP-ribosyl)ation (PARylation) is a post-translational protein modification catalyzed by enzymes known as poly(ADP-ribose) polymerases (PARPs). PARylation plays a pivotal role in regulating various cellular processes, including DNA repair, chromatin structure organization, gene expression, DNA replication, RNA processing, and ribosome biogenesis [1]. The primary enzymes catalyzing PARylation in the nucleus and synthesis of the elongated poly(ADP-ribose) polymer (PAR) are two DNA-dependent enzymes of the PARP family: PARP1 and PARP2 [2]. The functions of these enzymes in DNA repair regulation significantly overlap, as both participate in base excision repair (BER) and interact with key components of this process, such as DNA polymerase β, XRCC1, and APE1 [3]. PARP1 and PARP2 interact with DNA molecules mimicking intermediates of short- and long-patch BER pathways, as demonstrated using synthetic DNA duplexes containing lesions and nucleosomes [4]. Here, we elucidated differences in the interaction of PARP1 and PARP2 enzymes with DNA lesions – BER intermediates on model nucleosomes, as well as the influence of lesion position on histone modification by these enzymes.
Methods and Algorithms: Polyacrylamide gel electrophoresis, fluorescence spectrophotometry.
Results: Despite extensive studies on the affinity of PARP1 and PARP2 for various forms of DNA damage, including single-nucleotide gaps in the context of nucleosome core particles (NCPs), a detailed examination of the interaction of these enzymes with DNA lesions, depending on their distance from the 5'-end of DNA, has not been conducted. In this study, anisotropy fluorescence measurements were used to evaluate EC50 values reflecting the affinity of PARP1 and PARP2 for various DNA activators. We demonstrated that approaching the gap to the 5'-end of DNA led to increased EC50 values for PARP1, but not for PARP2. The decrease in affinity may be associated with the competition of PARP1 molecules for binding to two closely spaced lesions: the blunt end of DNA and a single-nucleotide gap. Meanwhile, PARP2 specifically interacts with

the gap with high affinity, and this damage is the sole site of specific interaction of PARP2 with NCP. Additionally, the amount and distribution of PAR synthesized by PARP1 and PARP2 in the presence of the histone PARylation factor HPF1 were analyzed. We showed that histone modification likely depends more on the position of the gap relative to the histone core than on the distance of the gap from the 5'-end of DNA.

Conclusion: The presence of a gap in the DNA structure within the nucleosome affects HPF1-dependent PARylation of histones. The affinity of PARP1 for a single-nucleotide gap in DNA depends on the distance of the gap from the 5'-end of DNA.

Funding: The study is supported by the grant from RFFS No. 22-74-10059.

Список литературы/References

1. Kraus W.L. PARPs and ADP-ribosylation: 60 years on. *Genes Dev*. 2020;34:251-253
2. Wei H., Yu X. Functions of PARylation in DNA damage repair pathways. *Genomics Proteomics Bioinf*. 2016;14:131-139
3. Vasil'eva I.A. et al. Dynamic light scattering study of base excision DNA repair proteins and their complexes. *Biochim Biophys Acta Proteins Proteom*. 2019;1867(3):297-305
4. Kutuzov M.M. et al. The contribution of PARP1, PARP2 and poly(ADP-ribosyl)ation to base excision repair in the nucleosomal context. *Sci Rep*. 2021;11:4849

doi 10.18699/bgrs2024-11.1-06

Особенности синтеза P-NH_2 олигонуклеотидов по адаптированной методике на основе стандартного амидофосфитного протокола

Малова Е.*, Пышная И., Мещанинова М., Пышный Д.

Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** malova.ev.an@gmail.com*

Ключевые слова: амидофосфатные производные олигодезоксирибонуклеотидов; автоматический синтез; реакция Штаудингера; оптимизация протокола синтеза

Мотивация и цель: В последние десятилетия молекулярная биология и медицинская химия нераздельно связаны с использованием синтетических олигонуклеотидов (ON) [1, 2]. Для изменения нуклеазоустойчивости или эффективности связывания с комплементарными ДНК или РНК особый интерес вызывают модифицированные по сахаро-фосфатному остову ON, в частности фосфат-модифицированные аналоги [3, 4]. Их широкое использование ограничивается возможностями стандартного автоматического твердофазного синтеза. Среди фосфат-модифицированных ON можно выделить N-незамещенные амидофосфатные аналоги, или P-NH_2 олигонуклеотиды. Являясь изоструктурными природным фосфодиэфирным нуклеиновым кислотам, они обладают способностью формировать устойчивые комплексы с комплементарной последовательностью ДНК/РНК при сниженной ионной силе раствора. Однако, планомерное использование P-NH_2 аналогов нуклеиновых кислот в качестве объектов и/или инструментов в молекулярной биологии и биомедицине ограничено сложностью их синтеза. На протяжение около 40 лет ведется поиск эффективных синтетических схем получения P-NH_2 олигонуклеотидов. В большинстве работ за основу взят H-фосфонатный протокол синтеза ON, включающий этап окислительного амидирования (напр., [5, 6]). Однако подходы, с использованием H-фосфонатной химии, не показали высоких выходов целевого продукта. Гораздо позднее был предложен метод постсинтетической модификации боранфосфонатных олигодезоксинуклеотидов посредством йод-зависимого замещения при атоме фосфора [7]. Подход, характеризуется высокими выходами, но основан на использовании специфической химии, в настоящее время коммерчески недоступной. Несмотря на количество работ, посвященных синтезу и исследованию свойств P-NH_2 олигонуклеотидов, дальнейшего развития данные аналоги не получили. Целью данной работы являлось создание адаптированной методики получения P-NH_2 аналогов, основанной на амидофосфитном протоколе синтеза ON, в качестве более простой и доступной альтернативы существующим подходам.

Методы и алгоритмы: Для синтеза P-NH_2 аналогов за основу был взят стандартный, широко используемый в настоящее время протокол амидофосфитного синтеза олигонуклеотидов. Введение P-NH_2 модификации осуществляли на стадии окисления с использованием азида, (9-флуоренил)-метоксикарбонилазида ($FmocN_3$), по реакции Штаудингера. Последующее

формирование N-незамещенного амидофосфатного фрагмента в составе олигонуклеотида осуществляли посредством удаления Fmoc-группы при обработке сильным основанием. Полученный по оптимизированной методике набор моно- и бис-модифицированных октамерных олигонуклеотидов использовали для исследования термодинамических свойств P-NH_2 аналогов в составе комплементарных НК-комплексов в условиях пониженной ионной силы раствора методом термической денатурации с оптической регистрацией сигнала.

Результаты: В ходе оптимизации методики первоначальное предположение о возможности постсинтетического формирования N-незамещенного амидофосфата было опровергнуто в связи с лабильностью промежуточного Fmoc-N-замещенного фрагмента в условиях синтетического цикла. В связи с чем, в разрабатываемый протокол была внедрена дополнительная стадия отщепления флуоренильной группы, следующая за каждой стадией окисления по реакции Штаудингера. Указанная последовательность стадий позволила достигнуть не менее 80 % конверсии для каждого цикла модификации. Отсутствие количественной конверсии в модифицированных фрагмент связано с долей неполного окисления по реакции Штаудингера. В настоящее время проводится поиск причин, препятствующих полному окислению фосфитного компонента органическим азидом. Продемонстрировано отсутствие явной зависимости выхода P-NH_2-олигонуклеотидов как от положения P-NH_2-звена в цепи, так и от нуклеозидного состава модифицируемого фрагмента. С помощью разработанного протокола синтезирован набор моно- и бис-модифицированных октадезоксирибонуклеотидов для детального исследования термической стабильности комплементарных ДНК/ДНК комплексов в различных буферных условиях. Значения температур плавления в условиях высокой ионной силы раствора (1 M NaCl, pH 7.2) согласуются с ранее полученными в литературе [6] и свидетельствуют о снижении термостабильности комплекса ДНК в среднем на 1.3 ºC при введении одного P-NH_2-звена. При уменьшении ионной силы раствора дестабилизирующий эффект P-NH2-модификации достоверно снижается, что дополнительно подтверждает электронейтральный статус вводимого амидофосфатного звена.

Выводы: Разрабатываемая методология для синтеза P-NH_2 аналогов олигонуклеотидов на основе стандартного амидофосфитного протокола синтеза олигонуклеотидов позволяет получать с хорошими выходами моно- и бис-модифицированные последовательности ON. Введение бóльшего числа модифицированных звеньев сопряжено с большими потерями олигонуклеотидного материала и требует дополнительной оптимизации синтетического процесса. Результаты исследования термической стабильности комплексов с P-NH_2 модифицированными цепями продемонстрировало потенциал использования данных аналогов в ситуациях подразумевающих применение растворов с пониженной ионной силой раствора.

Финансирование: Исследование поддержано Российским научным фондом (№ 21-64-00017).

Features of the synthesis of P-NH_2 oligonucleotides using an adapted method based on the standard phosphoramidite protocol

Malova E.*, Pyshnaya I., Meschaninova M., Pyshnyi D.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** malova.ev.an@gmail.com*

Key words: phosphoramidate derivatives of oligodeoxyribonucleotides; automated synthesis; Staudinger reaction; synthesis protocol optimization

Motivation and Aim: In recent decades, molecular biology and medicinal chemistry have been inextricably linked to the use of synthetic oligonucleotides (ON) [1, 2]. To modify nuclease stability or binding efficiency to complementary DNA or RNA, sugar-phosphate modified ON, in particular phosphate-modified analogs, are of particular interest [3, 4]. Their wide application is limited by the possibilities of standard automated solid-phase synthesis. Among phosphate-modified ON we can distinguish N-unsubstituted phosphoramidate analogs, or P-NH_2 oligonucleotides. Being isostructural with respect to natural phosphodiester nucleic acids, they have the ability to form stable complexes with complementary DNA/RNA sequence at reduced ionic strength of solution. However, the systematic use of P-NH_2-analogs of nucleic acids as objects and/or tools in molecular biology and biomedicine is limited by the complexity of their synthesis. The search for efficient synthetic schemes for the preparation of P-NH_2 oligonucleotides has been going on for about 40 years. In most studies, the H-phosphonate protocol for the synthesis of ON including an oxidative amidation step has been taken as a basis (e. g., [5, 6]). However, approaches using H-phosphonate chemistry did not show high yields of the target product. Much later, a method for the post-synthetic modification of borphosphonate oligodeoxynucleotides by iodine-dependent substitution at the phosphorus was proposed [7]. The approach is characterized by high yields, but is based on the use of specific chemistry, currently commercially unavailable. Despite the large number of works devoted to the synthesis and study of the properties of P-NH_2 oligonucleotides, these analogs have not been further developed. The aim of this work was to develop an adapted methodology for the preparation of P-NH_2 analogs based on the ON phosphoramidite synthesis protocol as a simpler and more affordable alternative to existing approaches.
Methods and Algorithms: For the synthesis of P-NH_2-analogs, a standard, currently widely used protocol for the phosphoramidite synthesis of oligonucleotides was taken as a basis. The introduction of the P-NH_2 modification was carried out in the oxidation step using the azide, (9-fluorenyl)-methoxycarbonylazide (FmocN_3), via the Staudinger reaction. Subsequent formation of an N-unsubstituted phosphoramidate moiety in the oligonucleotide was accomplished by removal of the Fmoc group by strong base treatment. A set of mono- and bis-modified octameric oligonucleotides obtained by the optimized method was used to study the thermodynamic properties of P-NH_2 analogues as part of complementary nucleic acid complexes under conditions of reduced ionic strength of solution by thermal denaturation with optical signal registration.
Results: During the optimization of the methodology, the initial assumption about the possibility of post-synthetic formation of N-unsubstituted phosphoramidate was refuted

due to the lability of the intermediate Fmoc-N-substituted fragment under synthetic cycle conditions. Therefore, an additional cleavage step of the fluorenyl group following each oxidative step of the Staudinger reaction was introduced into the developed protocol. The above sequence of steps allowed achieving a conversion of at least 80 % for each modification cycle. The lack of quantitative conversion in the modified fragments is due to the fraction of incomplete oxidation by the Staudinger reaction. The reasons preventing the complete oxidation of the phosphite component by the organic azide are searched for. It is shown that there is no obvious dependence of the yield of P-NH_2-oligonucleotides both on the position of the P-NH_2 link in the chain and on the nucleoside composition of the modified fragment. Using the developed protocol, a set of mono- and bis-modified octadeoxyribonucleotides was synthesized for a detailed study of the thermal stability of complementary DNA/DNA complexes under various buffer conditions. The values of melting temperatures under conditions of high ionic strength of the solution (1 M NaCl, pH 7.2) agree with those previously obtained in the literature [6] and indicate an average decrease in the thermostability of the DNA complex by 1.3 ºC upon introduction of a single P-NH_2 linkage. When the ionic strength of the solution is reduced, the destabilizing effect of the P-NH_2-modification decreases significantly, which further confirms the electroneutral status of the introduced phosphoramidate linkage.

Conclusion: The developed methodology for the synthesis of P-NH_2-analogs of oligonucleotides based on the standard phosphoramidite protocol for oligonucleotide synthesis allows to obtain mono- and bis-modified ON-sequences with good yields. The introduction of a larger number of modified links is associated with large losses of oligonucleotide material and requires additional optimization of the synthetic process. The results of the study of thermal stability of complexes with P-NH_2-modified chains showed the possibility of using these analogs in situations involving the use of solutions with reduced ionic strength of the solution.

Funding: The study is supported by Russian Science Foundation (No. 21-64-00017).

Список литературы/References

1. Egli M., Manoharan M. Chemistry, structure and function of approved oligonucleotide therapeutics. *Nucleic Acids Res*. 2023;51(6):2529-2573
2. Green M.R., Sambrook J. Polymerase Chain Reaction. *Cold Spring Harb Protoc*. 2019;2019(6):436-456
3. Gołębiewska J., Stawinski J. Boranephosphonates. Unraveling chemistry of the P-BH_3 functional group. *Phosphorus Sulfur Silicon Relat Elem*. 2021;197(5-6):480-486. doi 10.1080/10426507.2021.1990922
4. Stetsenko D., Pyshnyi D., Kupryushkin M. Modified oligonucleotides and methods for their synthesis. Patent WO2016028187A1. 2016
5. Iyer R.P., Devlin T., Habus I., Yu D., Johnson S., Agrawal S. Oligonucleoside phosphoramidates from N-pent-4-enoyl nucleoside H-phosphonates. *Tetrahedron Lett*. 1996;37(10):1543-1546
6. Peyrottes S., Vasseur J.J., Imbach J.L., Rayner B. Replacement of the phosphodiester backbone of oligodeoxynucleotide analogues by non-ionic phosphoramidate (P-NH2). *Nucleosides Nucleotides*. 1997;16(7-9):1551-1554
7. Paul S., Roy S., Monfregola L., Shang S., Shoemaker R., Caruthers M.H. Oxidative substitution of boranephosphonate diesters as a route to post-synthetically modified DNA. *J Am Chem Soc*. 2015;137(9):3253-3264

doi 10.18699/bgrs2024-11.1-07

Апуриновая/апиримидиновая эндонуклеаза 1 и тирозил-ДНК-фосфодиэстераза 1 препятствуют накоплению необратимых ДНК-белковых комплексов

Речкунова Н.*, Лебедева Н., Ендуткин А., Дырхеева Н., Лаврик О.

Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** nadyarec@niboch.nsc.ru*

Ключевые слова: ДНК-белковые сшивки; апуриновые/апиримидиновые сайты; апуриновая/апиримидиновая эндонуклеаза 1; тирозил-ДНК-фосфодиэстераза 1; репарация ДНК

Мотивация и цель: ДНК-белковые сшивки (DPC) – распространенные повреждения ДНК, вызываемые различными внутриклеточными и внешними факторами. Такие повреждения очень токсичны, поскольку из-за большого размера блокируют процессы, протекающие в хроматине, и не способны удаляться каноническими системами репарации ДНК [1]. Один из главных источников таких повреждений – апуриновые/апиримидиновые (AP) сайты, образующие сшивки с белками. Целью работы было исследование роли ферментов, осуществляющих процессинг AP-сайтов – апуриновой/апиримидиновой эндонуклеазы 1 (APE1) и тирозил-ДНК-фосфодиэстеразы 1 (TDP1) – в предотвращение или удаление ДНК-белковых аддуктов, формируемых в отсутствие восстанавливающих агентов. Такие аддукты считаются суицидальными, поскольку могут образовываться в клетке и стабильны без дополнительной обработки.

Методы и алгоритмы: В работе использовали очищенные рекомбинантные белки, а также экстракты клеток, полученных из линий HEK293FT дикого типа и нокаутных по генам (APE1$^{-/-}$) или TDP1 (TDP1$^{-/-}$). AP-сайт в составе ДНК генерировали путем обработки ДНК-дуплекса, содержащего урацил или 8-охогуанин, соответствующей ДНК-гликозилазой (UDG или OGG1). Формирование сшивок белков с радиоактивно меченными AP-сайт-содержащими ДНК-дуплексами анализировали электрофорезом в ПАА геле без предварительной обработки проб $NaBH_4$ с последующей радиоавтографией.

Результаты: Ранее было показано, что белки-участники процесса эксцизионной репарации оснований (BER) – ДНК-гликозилаза OGG1, поли(ADP-рибоза) полимеразы 1 (PARP1) и 2 (PARP2), а также белок-платформа XRCC1 – взаимодействуют с AP-сайтом с образованием основания Шиффа, после восстановления которого с помощью $NaBH_4$ формируются стабильные ковалентные аддукты [2]. Кроме того, было показано, что PARP1 образует стабильный ковалентный комплекс с ДНК, содержащей AP-сайт (AP-ДНК), без восстановления $NaBH_4$ [3]. Мы исследовали, является ли образование стабильной ДНК-белковой сшивки (DPC) в отсутствие $NaBH_4$ уникальной особенностью PARP1, или другие белки, взаимодействующие с AP-сайтом, также способны образовывать такие аддукты [4]. Проведенный анализ показал, что не только PARP1, но и OGG1, PARP2 и XRCC1 образуют стабильные сшивки с AP-ДНК без восстановления $NaBH_4$.

Ключевой фермент BER, расщепляющий AP-сайты, – APE1. Расщепление AP-сайта также может катализировать TDP1, хотя и с меньшей эффективностью, чем APE1 [5]. Мы показали, что гидролиз AP-сайта с помощью APE1 предотвращает

образование DPC, тогда как TDP1 способен расщеплять DPC, удаляя из ковалентного комплекса с ДНК некоторые белки, в частности OGG1 и PARP2. В то же время TDP1 не способен расщеплять DPC, образованные высокомолекулярными белками, в том числе PARP1, а также XRCC1, который незначительно превышает по молекулярной массе TDP1, но имеет неглобулярную структуру. Мы предполагаем, что TDP1 удаляет белки, присоединенные к ДНК путем ковалентной сшивки с 3'-фосфо-α,β-ненасыщенным альдегидом, образующимся в результате расщепления АР-сайта под действием АР-лиазной активности OGG1 или других белков, по тому же механизму, что и другие аддукты с 3'-фосфатной группы ДНК [5].

Мы также исследовали влияние APE1 и TDP1 на эффективность образования DPC в экстрактах клеток APE1$^{-/-}$ или TDP1$^{-/-}$. Наши данные показывают, что нокаут гена, кодирующего APE1, приводит к повышению уровня сшивок PARP1-ДНК в клеточном экстракте, тогда как дефицит TDP1 мало влияет на образование DPC.

Выводы: Таким образом, мы продемонстрировали, что APE1 играет решающую роль в предотвращении образования DPC в клеточных экстрактах. Напротив, TDP1 удаляет некоторые белковые аддукты с ДНК, используя тот же механизм, что и для удаления других групп, связанных с 3'-фосфатом ДНК, но не способен предотвращать их образование из-за низкой эффективности расщепления АР-сайта по сравнению с APE1. Поскольку активность APE1 остается высокой в клетках TDP1$^{-/-}$, уровни DPC в экстрактах из мутантных клеток и клеток дикого типа были очень похожи и ниже, чем в экстракте клеток APE1$^{-/-}$.

Финансирование: Исследование поддержано Российским научным фондом (№ 21-64-00017).

Apurinic/apyrimidinic endonuclease 1 and tyrosyl-DNA phosphodiesterase 1 prevent suicidal covalent DNA-protein crosslinks

Rechkunova N.*, Lebedeva N., Endutkin A., Dyrkheeva N., Lavrik O.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** nadyarec@niboch.nsc.ru*

Key words: DNA-protein crosslinks; apurinic/apyrimidinic site; apurinic/apyrimidinic endonuclease 1; tyrosyl-DNA phosphodiesterase 1; DNA repair

Motivation and Aim: DNA-protein crosslinks (DPC) are common DNA lesions induced by various external and endogenous agents. Such lesions are highly toxic due to the bulk, resulting in virtual blocking of chromatin-based processes and their inability to be repaired by canonical repair pathways [1]. Apurinic/apyrimidinic sites (AP sites) and proteins interacting with them are one of the main sources of DPC. The aim of the work was to investigate the role of AP site processing enzymes, apurinic/apyrimidinic endonuclease 1 (APE1) and tyrosyl-DNA phosphodiesterase 1 (TDP1), in preventing or removing DNA-protein adducts formed in the absence of reducing agents. Such DNA-protein crosslinks can arise in cells where there are no additional reducing agents.

Methods and Algorithms: Purified recombinant proteins, as well as extracts of cells obtained from wild-type HEK293FT and APE1 or TDP1 knockout lines (APE1$^{-/-}$ or TDP1$^{-/-}$) were used. The AP site in DNA was generated by treating a DNA duplex containing uracil or 8-oxoguanine with the appropriate DNA glycosylase (UDG or

OGG1). The formation of protein crosslinks with radioactively labeled AP site-containing DNA duplexes was analyzed by electrophoresis in a PAA gel without treatment of the samples with $NaBH_4$, followed by autoradiography.

Results: It was previously shown that proteins involved in the base excision repair (BER) process – DNA glycosylase OGG1, poly(ADP-ribose) polymerases 1 (PARP1) and 2 (PARP2), as well as the scaffold protein XRCC1 – interact with the AP site with the formation of a Schiff base, after reduction of which with $NaBH_4$ stable covalent adducts are formed [2]. In addition, PARP1 has been shown to form a stable covalent complex with DNA containing an AP site (AP-DNA) without $NaBH_4$ reduction [3]. We investigated whether the formation of a stable DNA-protein crosslinks (DPC) in the absence of $NaBH_4$ is a unique feature of PARP1 or whether other AP site-interacting proteins are also capable to form such adducts [4]. The analysis showed that not only PARP1, but also OGG1, PARP2 and XRCC1 form stable crosslinks with AP-DNA without $NaBH_4$ reduction.

The key BER enzyme that cleaves AP sites is APE1. AP site cleavage can also be catalyzed by TDP1, although with less efficiency than APE1 [5]. We have shown that hydrolysis of the AP site by APE1 prevents the formation of DPC, while TDP1 is able to resolve DPC, removing some proteins, in particular OGG1 and PARP2, from the covalent complex with DNA. At the same time, TDP1 is not able to cleave DPCs formed by high molecular weight proteins, including PARP1, as well as XRCC1, which slightly exceeds the molecular weight of TDP1, but has a non-globular structure. We propose that TDP1 removes proteins attached to DNA by covalent crosslinking with a 3'-phospho-α,β-unsaturated aldehyde resulting from AP site cleavage by the AP lyase activity of OGG1 or other proteins by the same mechanism as other adducts with the 3'-phosphate group of DNA [5].

We also examined the effect of APE1 and TDP1 on the efficiency of DPC formation in extracts of $APE1^{-/-}$ or $TDP1^{-/-}$ cells. Our data show that knockout of the gene encoding APE1 leads to increased level of PARP1-DNA crosslinks in the cell extract, whereas TDP1 deficiency has little effect on DPC formation.

Conclusion: In summary, we demonstrated that APE1 plays a critical role in preventing DPC formation in cell extracts. In contrast, TDP1 resolves some protein adducts with DNA using the same mechanism as for the removal of other groups bound to the DNA 3′-phosphate, but is unable to prevent their formation due to low AP site cleavage efficiency compared to APE1. Because APE1 activity remains high in $TDP1^{-/-}$ cells, DPC levels in extracts from mutant and wild-type cells were very similar and lower than in extract from $APE1^{-/-}$ cells.

Funding: The study is supported by the Russian Science Foundation (No. 21-64-00017).

Список литературы/References

1. Weickert P., Stingele J. DNA-Protein Crosslinks and Their Resolution. *Annu Rev Biochem*. 2022;91:157-181
2. Khodyreva S., Lavrik O. Non-canonical interaction of DNA repair proteins with intact and cleaved AP sites. *DNA Repair*. 2020;90:102847
3. Prasad R. et al. Suicidal cross-linking of PARP-1 to AP site intermediates in cells undergoing base excision repair. *Nucleic Acids Res.* 2014;42:6337-6351
4. Lebedeva N.A. et al. Apurinic/Apyrimidinic Endonuclease 1 and Tyrosyl-DNA Phosphodiesterase 1 Prevent Suicidal Covalent DNA-Protein Crosslink at Apurinic/Apyrimidinic Site. *Front Cell Dev Biol*. 2021;8:617301
5. Interthal H. et al. Human TDP1 cleaves a broad spectrum of substrates, including phosphoamide linkages. *J Biol Chem*. 2005;280:36518-36528

doi 10.18699/bgrs2024-11.1-08

HPF1 регулирует поли(ADP-рибозил)ирование РНК-связывающего белка FUS и влияет на образование поли(ADP-рибоза)-содержащих компартментов

Сингатулина А.*, Суханова М., Лаврик О.

Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия

** nastsing@yandex.ru*

Ключевые слова: репарация ДНК; поли(ADP-рибоза)полимеразы 1 и 2 (PARP1/2); фактор поли(ADP-рибозил)ирования гистонов 1 (HPF1)

Мотивация и цель: В настоящее время многие РНК-связывающие белки рассматриваются как участники поддержания стабильности генома, поскольку способны напрямую участвовать в репарации ДНК и могут локализоваться вблизи сайтов повреждений ДНК за счет их взаимодействия с поли(ADP-рибозой) [1]. Поли(ADP-рибоза) (PAR) – это отрицательно заряженный разветвленный полимер, состоящий из мономеров ADP-рибозы. Синтез PAR из NAD^+ катализируют ферменты поли(ADP-рибоза)полимеразы (PARPs), которые активируются при взаимодействии с поврежденной ДНК [2]. Ядерные ферменты PARP1 и PARP2 катализируют автополи(ADP-рибозил)ирование (авто-PAR-илирование), а также PAR-илирование других белков-мишеней. Недавно был открыт фактор PAR-илирования гистонов 1, белок HPF1, который способен образовывать временный совместный активный центр с PARP1(2), изменять специфичность реакции PAR-илирования и регулировать уровень синтеза PAR. В данной работе мы исследовали влияние HPF1 на активность PARP1(2) и формирование компартментов с участием PAR-илированной PARP1(2) и РНК-связывающего белка FUS (FUS/TLS, Fused in Sarcoma). FUS-зависимое образование компартментов представляет большой интерес для исследования, поскольку FUS взаимодействует с PAR, а также способен локализоваться в сайтах повреждения ДНК, и этот процесс зависит от активности PARPs и синтеза PAR [3]. Кроме того, ранее методом атомно-силовой микроскопии (АСМ) нами было показано, что FUS и PAR-илированный PARP1 образуют сложные надмолекулярные структуры (компартменты), в которых концентрируется поврежденная ДНК [4].

Методы и алгоритмы: АСМ: в образцы добавляли путресцин до конечной концентрации 5 мМ и магний (5 мМ) и наносили на поверхность свежесколотой слюды и инкубировали в течение 1 минуты. Поверхность слюды промывали 0.02 % раствором уранилацетата с последующей промывкой водой и высушивали на воздухе. АСМ-исследования проводили в полуконтактном режиме на микроскопе Nanoscope V Multimode 8 на воздухе с использованием кантиливера Scanasyst-Air. Электрофореграмма PAR-илированных белков по Леммли: анализ PAR-илирования белков проводили методом денатурирующего гель-электрофореза. Гели визуализировали с помощью сканера Typhoon FLA 9500 или окрашиванием коллоидным Кумасси G250.

Результаты: Важную роль в формировании компартментов играет взаимодействие FUS с PAR [4], поэтому структура PAR, ее длина и уровень синтеза могут оказывать существенное влияние на этот процесс. HPF1 регулирует активность PARP1(2), оказывая влияние на уровень автомодификации этих белков и длину синтезируемой PAR. С помощью биохимических методов мы показали, что с увеличением концентрации HPF1 уменьшается размер синтезируемого PAR в реакции автомодификации PARP1(2). Для анализа размера и количества компартментов, образующихся с участием FUS после PAR-илирования PARP1(2) в отсутствие или в присутствии различных концентраций HPF1, мы провели серию экспериментов с использованием АСМ. Мы обнаружили, что во время автомодификации PARP1 и PARP2 в присутствии HPF1 размер PAR-илированного белка уменьшается. Кроме того, FUS способен образовывать компартменты с участием PAR-илированного PARP2.). С помощью АСМ мы также наблюдали, что увеличение концентрации HPF1 приводило к уменьшению размеров компартментов, образующихся при взаимодействии FUS с PAR-илированным PARP1(2), но при этом количество компартментов оставалось на прежнем уровне (рис. 1).

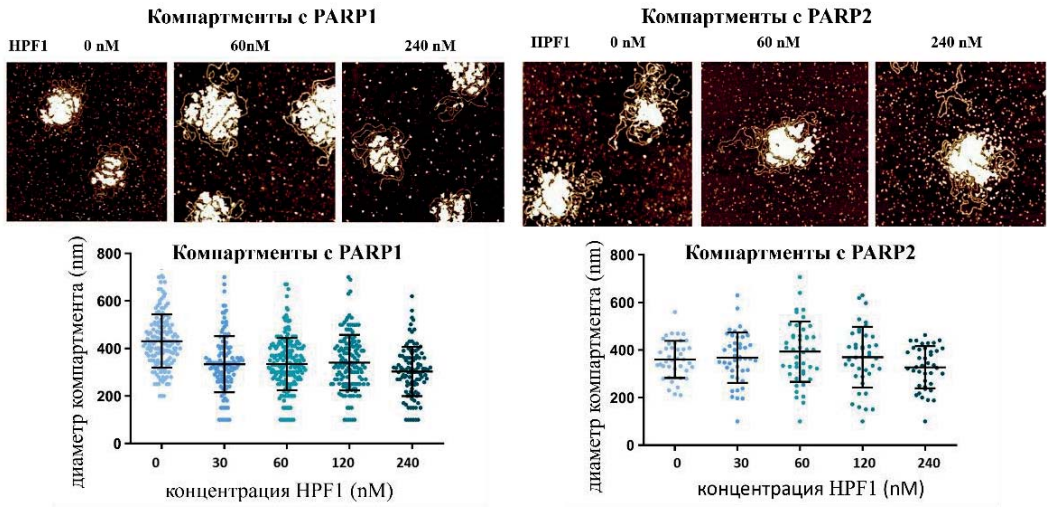


Рис. 1. Визуализация компартментов, сформированных с участием FUS после авто-PAR-илирования PARP1(2) в присутствии различной концентрации HPF1 (верхние панели: АСМ-изображения компартментов). Анализ размеров компартментов методом АСМ (нижние графики)

Ранее было показано, что FUS не только способен взаимодействовать с PAR, но и является мишенью в реакции PAR-илирования, катализируемой PARP1. Согласно литературным данным, в присутствии HPF1 специфичность действия PARP1(2) изменяется, и вместо аминокислотных остатков аспартата и глутамата, как это происходит в отсутствие HPF1, основными сайтами модификации становятся остатки серина. Поэтому возможно, что присутствие HPF1 может оказывать влияние на PAR-илирование FUS за счет активности PARP1(2).

Далее была проведена серия экспериментов по изучению влияния HPF1 на PAR-илирование FUS (WT) и его фосфомиметика (FUS 12E), содержащего замены остатков серина/треонина на глутамин, или мутантных форм, содержащих делецию N-концевого (FUS ΔLCD) или двух С-концевых (FUS ΔRGG1-2) доменов (рис. 2). Было показано, что в присутствии HPF1 происходит значительная стимуляция PAR-илирования FUS и его мутантов за счет активности PARP1 и PARP2. Важно отметить, что HPF1 оказывал более сильный стимулирующий эффект на уровень PAR-илирования FUS белком PARP2. Изучение PAR-илирования FUS в присутствии HPF1 важно, поскольку FUS является неотъемлемым компонентом PAR-содержащих компартментов и принимает непосредственное участие в их формировании.

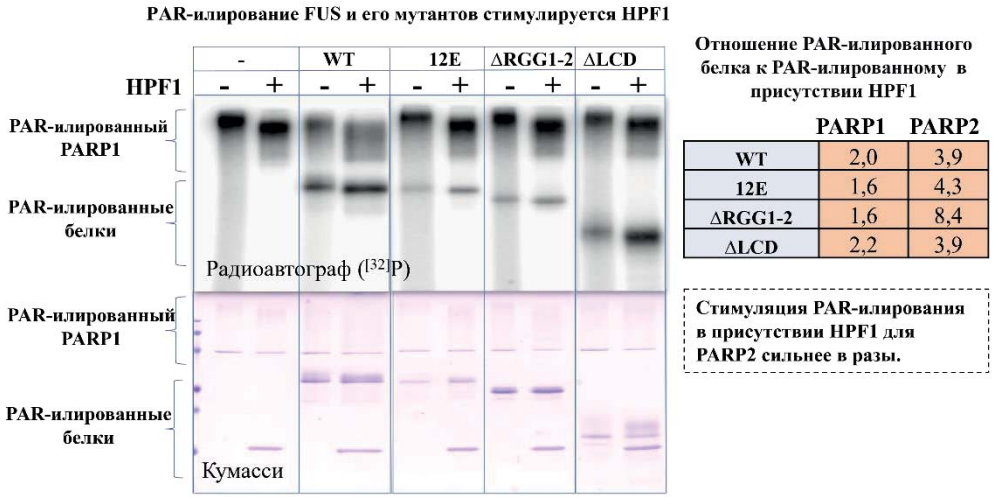


Рис. 2. HPF1 значительно увеличивает уровень PAR-илирования белка FUS (WT) и его мутантов. PAR-илирование белка FUS и его мутантов (12E, ΔLCD ΔRGG1-2) ферментами PARP1 и PARP2 в присутствии HPF1 анализировали с использованием радиоактивно меченного NAD^+ по уровню включения [^{32}P]-меченной ADP

Выводы: Таким образом, HPF1 оказывает ингибирующее действие на уровень авто-PAR-илирования PARP1 и PARP2, тем самым оказывая влияние на размер компартментов, содержащих FUS, PAR-илированный белок и поврежденную ДНК. Однако в присутствии HPF1 происходит значительная стимуляция PAR-илирования самого белка FUS и его мутантов. Полученные результаты указывают на возможную роль HPF1 в регуляции формирования FUS-PAR-зависимых компартментов в процессе репарации ДНК. Это может иметь важное значение для разработки и использования ингибиторов PARP1(2) в противоопухолевой терапии.

Финансирование: Исследование поддержано грантом РНФ (№ 20-14-00086).

HPF1 regulates poly(ADP-ribosyl)ation of the RNA-binding protein FUS and influences the formation of poly(ADP-ribose)-containing compartments

Singatulina A.*, Sukhanova M., Lavrik O.

Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia

** nastsing@yandex.ru*

Key words: DNA repair; poly(ADP-ribose) polymerases 1 and 2 (PARP1/2); histone poly(ADP-ribosyl)ation factor 1 (HPF1)

Motivation and Aim: Currently, many RNA-binding proteins are considered to be involved in maintaining genome stability, since they are able to directly participate in DNA repair and can be localized at the sites of DNA damage due to their ability to interact with poly(ADP-ribose) [1]. Poly(ADP-ribose) (PAR) is a negatively charged branched polymer consisting of ADP-ribose monomers. The synthesis of PAR from NAD^+ is catalyzed by the enzymes poly(ADP-ribose) polymerases (PARPs), which are activated upon interaction with damaged DNA [2]. The nuclear enzymes PARP1 and PARP2 can catalyze autopoly(ADP-ribosyl)ation (auto-PARylation), as well as PARylation of other target proteins. Recently, the histone PARylation factor 1 (HPF1) was discovered, it is capable to form a temporary joint active site with PARP1(2), changing the specificity of the PARylation reaction and regulating the level of PAR synthesis. In this work, we investigated the effect of HPF1 on the activity of PARP1(2) and the formation of compartments involving PARylated PARP1(2) and the RNA-binding protein FUS (FUS/TLS, Fused in Sarcoma). FUS-dependent formation of compartments is of great interest for research, since FUS interacts with PAR and is also able to localize at the sites of DNA damage, and this process depends on the activity of PAR synthesis [3]. In addition, we previously showed using atomic force microscopy (AFM) that FUS and PARylated PARP1 form complex supramolecular structures (compartments) in which damaged DNA is concentrated [4].

Methods and Algorithms: AFM: putrescine was added to the samples to a final concentration of 5 mM and magnesium (5 mM) and applied to the surface of freshly cleaved mica and incubated for 1 minute. The mica surface was washed with a 0.02 % solution of uranyl acetate, followed by washing with water and dried in air. AFM studies were carried out in semi-contact mode on a Nanoscope V Multimode 8 microscope in air using a Scanasyst-Air cantilever. Electropherogram of PARylated proteins according to Laemmli: analysis of PARylated proteins was carried out by denaturing gel electrophoresis. Gels were visualized using a Typhoon FLA 9500 scanner or colloidal Coomassie G250 staining.

Results: The interaction of FUS with PAR plays an important role in the formation of compartments [4], therefore the structure of PAR, its length and level of synthesis can have a significant impact on this process. HPF1 regulates the activity of PARP1(2), influencing the level of auto-modification of these proteins and the length of synthesized PAR. Using biochemical methods, we showed that the size of PAR synthesized in the PARP1(2) automodification reaction was decreased with increasing of HPF1 concentration. To analyze the size and number of compartments formed by FUS after PARylation of PARP1(2) in the presence of various concentrations of HPF1, we

performed a series of experiments using AFM. We found that the size of the PARylated protein was decreased during automodification of PARP1 and PARP2 in the presence of HPF1. In addition, FUS is able to form compartments involving PARylated PARP2. Using AFM, we also observed that an increase in the concentration of HPF1 led to a decrease in the size of compartments formed during the interaction of FUS with PARylated PARP1(2), but the number of compartments remained at the same level (Fig. 1).

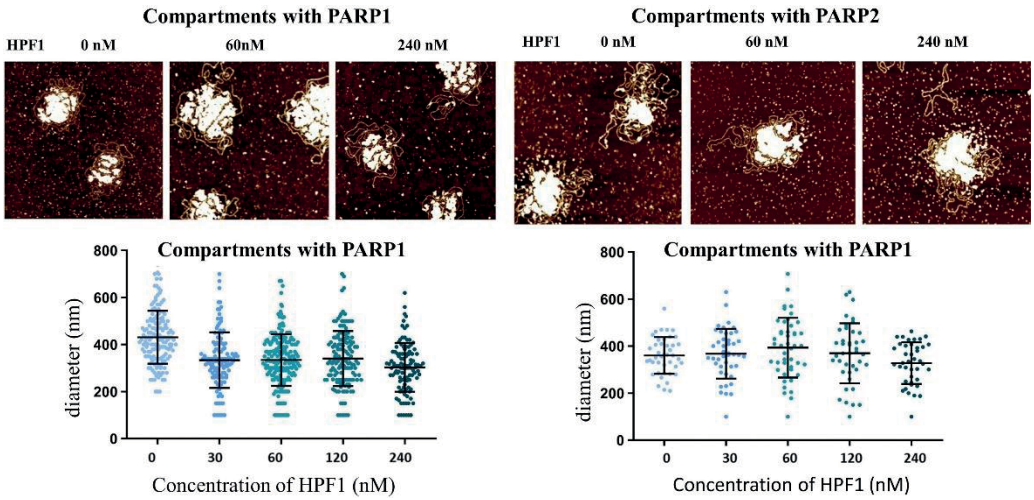
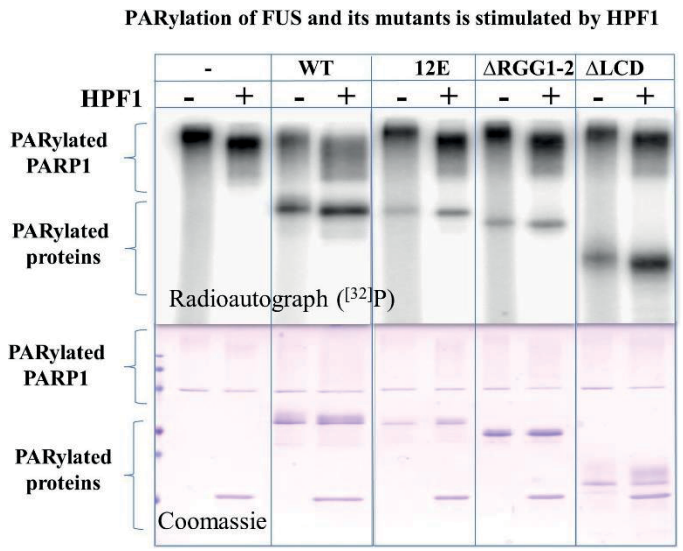


Fig. 1. Visualization of compartments formed with the participation of FUS after auto-PARylation of PARP1(2) in the presence of different concentrations of HPF1 (top panels: AFM images of compartments). Analysis of compartment sizes by AFM (lower graphs)

It was previously shown that FUS is not only able to interact with PAR, but it is also a target in the PARylation reaction catalyzed by PARP1. According to the literature, in the presence of HPF1, the specificity of PARP1(2) action changes, and instead of the amino acid residues of aspartate and glutamate, as happens in the absence of HPF1, serine residues become the main modification sites. It is therefore possible that the presence of HPF1 may influence FUS PARylation through PARP1(2) activity.

Next, a series of experiments was carried out to study the effect of HPF1 on PARylation of FUS (WT) and its phosphomimetic (FUS 12E), containing replacements of serine/threonine residues with glutamine, or mutant forms containing a deletion of the N-terminal (FUS ΔLCD) or two C-terminal (FUS ΔRGG1-2) domains (Fig. 2). It has been shown that in the presence of HPF1 there is a significant stimulation of PARylation of FUS and its mutants due to the activity of PARP1 and PARP2. Importantly, HPF1 had a stronger stimulatory effect on the level of FUS PARylation by PARP2. Studying FUS PARylation in the presence of HPF1 is important because FUS is an integral component of PAR-containing compartments and is directly involved in their formation.



Ratio of PARylated protein to PARylated in the presence of HPF1

	PARP1	PARP2
WT	2,0	3,9
12E	1,6	4,3
ΔRGG1-2	1,6	8,4
ΔLCD	2,2	3,9

Stimulation of PARylation in the presence of HPF1 for PARP2 is several times stronger.

Fig. 2. HPF1 significantly increases the level of PARylation of the FUS protein (WT) and its mutants. PARylation of the FUS protein and its mutants (12E, ΔLCD ΔRGG1-2) by the enzymes PARP1 and PARP2 in the presence of HPF1 was analyzed using [^{32}P]-labelled NAD^+ based on the level of incorporation of [^{32}P]-ADP-ribose

Conclusion: Thus, HPF1 has an inhibitory effect on the level of auto-PARylation of PARP1 and PARP2, thereby influencing the size of compartments containing FUS, PARylated protein and damaged DNA. However, there is a significant stimulation of PARylation of FUS and its mutant forms in the presence of HPF1. The results indicate a possible role of HPF1 in regulating the formation of FUS-PAR-dependent compartments during DNA repair. This may have important implications for the development and using of PARP1(2) inhibitors in anticancer therapy.

Funding: The study is supported by a grant from the Russian Science Foundation (No. 20-14-00086).

Список литературы/References

1. Bock F.J., Todorova T.T., Chang P. RNA regulation by poly (ADP-ribose) polymerases. *Mol Cell*. 2015;58(6):959-969
2. Alemasova E.E., Lavrik O.I. A sePARate phase? Poly (ADP-ribose) versus RNA in the organization of biomolecular condensates. *Nucleic Acids Res*. 2022;50(19):10817-10838
3. Rulten S.L. et al. PARP-1 dependent recruitment of the amyotrophic lateral sclerosis-associated protein FUS/TLS to sites of oxidative DNA damage. *Nucleic Acids Res*. 2014;2(1):307-314
4. Singatulina A.S. et al. PARP-1 activation directs FUS to DNA damage sites to form PARG-reversible compartments enriched in damaged DNA. *Cell Rep*. 2019;27:1809-1821

doi 10.18699/bgrs2024-11.1-09

Каталитическая активность REV1 человека

Столяренко А.[1, 2]*, Новикова А.[1, 2], Шилкин Е.[1, 2], Кручинин А.[1, 2], Макарова А.[1, 2]*

[1] *Национальный исследовательский центр «Курчатовский институт», Москва, Россия*

[2] *Институт биологии гена РАН, Москва, Россия*

* *nanostol@gmail.com, amakarova-img@yandex.ru*

Key words: REV1; ДНК-полимераза; транслезионный синтез ДНК; внутрицепочечная цисплатиновая сшивка; Хугстиновские взаимодействия

Мотивация и цель: Повреждения ДНК могут вызвать арест репликации, что губительно для живой клетки. Транслезионный синтез ДНК представляет собой механизм репликации повреждений ДНК, предотвращающий арест репликации. Консервативный эукариотический белок REV1 служит плацдармом для координации ДНК-полимераз во время транслезионного синтеза ДНК [1, 2]. Помимо структурной роли REV1 представляет собой ДНК-полимеразу Y-семейства с собственной дистрибутивной дезоксицитидилтрансферазной активностью [3]. Каталитическая активность обеспечивается уникальным механизмом с участием матрицы белковой природы, когда входящий dCTP и основание ДНК не взаимодействуют друг с другом напрямую [4]. Изменения экспрессии и/или каталитической активности REV1 способствуют мутагенезу и канцерогенезу и могут модулировать эффект химиотерапевтического лечения [5]. Известно, что REV1 *in vivo* участвует в репликации АП-сайтов [6-8], 1,N^6-ethenoA [7, 9], G-квадруплексов [10] и играет роль при соматической гипермутации генов иммуноглобулинов [11], однако, данные о точности и эффективности синтеза ДНК REV1 в литературе противоречивы. Целью работы является расширение знаний об активности REV1 на ДНК с разными повреждениями *in vitro*.

Методы и алгоритмы: Мы экспрессировали и выделили полноразмерный белок REV1 человека из эукариотической экспрессионной системы на основе *Saccharomyces cerevisiae*. Мы охарактеризовали его ДНК-полимеразную активность на неповрежденной ДНК и широком спектре поврежденных ДНК-субстратов. Отсутствие примесей, влияющих на активность REV1, проверяли с использованием каталитически неактивного варианта REV1. ДНК-полимеразы Pol iota и Pol eta человека были использованы в качестве контроля для анализа Хугстиновских и Уотсон-Криковских взаимодействий соответственно.

Результаты: G является предпочтительной матрицей REV1 [3, 12–15]. Мы проанализировали влияние ряда биологических и искусственных повреждений матричного G. Мы показали, что REV1 блокируется внутрицепочечной цисплатиновой сшивкой 1,2-GG, преобладающим повреждением, вызываемым химиотерапевтическим препаратом цисплатином. Искусственное повреждение, 7-deazaG, используемое для изучения Хугстиновских взаимодействий, также ингибировало REV1, что объясняется образованием контактов Хугстиновской стороны матричного нуклеотида с белком. Также мы протестировали репликацию наиболее распространенного повреждения G – 8-охоG (с противоречивыми данными в литературе [3, 13, 16, 17]) и показали, что REV1 ингибируется более чем в ~360 раз. Одновременно с этим эффект O^6-meG (~36-кратное ингибирование) соответствовал литературным данным [3, 18].

A является второй предпочтительной матрицей REV1 после G. Используя повреждение Уотсон-Криковской плоскости 1,N^6-ethenoA и модификацию Хугстиновской плоскости 7-deazaA, мы показали, что REV1 в гораздо большей степени ингибируется повреждением Хугстиновской плоскости. Данные литературы для 1,N^6-ethenoA противоречивы [13, 19]. Мы показали, что REV1 активируется на матрице с 1,N^6-ethenoA более чем в 50 раз, в соответствии с [19]. Кроме того, мы впервые проанализировали активность напротив 8-охоA и показали, что это повреждение легко реплицируется REV1 в ошибочной манере.

Среди модифицированных пиримидинов, при том что неповрежденные пиримидины являются наименее предпочтительными матрицами REV1, метилцитозин (mC) и гидроксиметилцитозин C (hmC) несколько ингибировали REV1. Тимингликоль (Tg) неожиданно реплицировался очень эффективно: более чем в 1700 раз эффективнее, чем неповрежденный T, и в ошибочной манере. Следует отметить, что для большинства исследованных повреждений (8-охоG, O^6-meG, 1,N^6-ethenoA, Tg) дискриминация нуклеотидных субстратов происходила на этапе связывания, что сопровождалось значительным изменением K_M, но не k_{cat}.

Выводы: В данной работе впервые была изучена активность REV1 человека на ДНК с внутрицепочечной цисплатиновой сшивкой 1,2-GG, с аналогами нуклеотидов 7-deazaG и 7-deazaA, с вызванными окислительным стрессом повреждениями 8-охоA и Tg, а также эпигенетическими модификациями mC и hmC.

Структуры REV1 с матрицами A, T и C в литературе не описаны. Мы приводим биохимические доказательства важности Хугстиновой плоскости матричного A для работы REV1. Известно, что матричный G контактирует атомами N7 и O6 своей Хугстиновской плоскости с амидами основной цепи His774 и Gly775 активного центра REV1 [4]. Наши данные указывают на то, что связывание матриц G и A в активном сайте REV1 происходит по сходному механизму (A находится в связывающем кармане фермента в той же конформации, что и G).

REV1 «активируется» на ДНК с 1,N^6-ethenoA и Tg. Интактные матричные A, T и C являются гораздо менее предпочтительными субстратами для REV1, поскольку они не способны образовывать взаимодействия, специфичные для матричного G, где атом O6 является акцептором водородной связи. Повышение активности REV1 напротив 1,N^6-ethenoA может объясняться потерей атомом в положении 6 свойств донора водорода [19]. Более того, согласно нашим результатам и литературным данным, присоединение CH3-группы к C с образованием mC или к U с образованием T [13, 20, 21] снижает активность REV1. Однако добавление OH-группы непосредственно к этой группе (к mC с образованием hmC) или вблизи нее (к T с образованием Tg) восстанавливает или даже усиливает активность REV1. Вместе эти данные проливают свет на функционирование активного центра REV1 на различных ДНК-субстратах.

Результаты частично опубликованы в работе https://www.mdpi.com/1422-0067/25/7/4107 (Stolyarenko A., Novikova A., Shilkin E., Poltorachenko V., Makarova A. The Catalytic Activity of Human REV1 on Undamaged and Damaged DNA. *International Journal of Molecular Sciences*. 2024;25:4107–4120. doi 10.3390/ijms25074107).

Финансирование: Исследование поддержано госзаданием НИЦ «Курчатовский институт».

Human REV1 catalytic activity

Stolyarenko A.[1, 2]*, Novikova A.[1, 2], Shilkin E.[1, 2], Kruchinin A.[1, 2], Makarova A.[1, 2]*

[1] *National Research Center "Kurchatov Institute", Moscow, Russia*

[2] *Institute of Gene Biology of the Russian Academy of Sciences, Moscow, Russia*

** nanostol@gmail.com, amakarova-img@yandex.ru*

Key words: REV1; DNA polymerase; translesion DNA synthesis; cisplatin intrastrand crosslink; Hoogsteen interactions

Motivation and Aim: DNA damage can cause replication arrest which is detrimental to the living cell. DNA translesion synthesis is a mechanism to bypass DNA lesions preventing replication arrest. Conserved eukaryotic REV1 protein serves as a scaffold for the coordination of DNA polymerases during DNA translesion synthesis [1, 2]. Besides its structural role, REV1 is a Y-family DNA polymerase with its own distributive deoxycytidyl transferase activity [3]. It is driven by a unique protein-template mechanism, when the incoming dCTP and the template base do not directly pair with each other [4]. Alterations in REV1 expression and/or catalytic activity promote mutagenesis and carcinogenesis and can modulate the effect of chemotherapy treatment [5]. Although REV1 activity is known to function *in vivo* in the bypass of AP sites [6-8], 1,N^6-ethenoA [7, 9], G quadruplexes [10] and in somatic hypermutation of immunoglobulin genes [11], data about the accuracy and efficiency of DNA synthesis by REV1 in the literature are contrasting. The aim of this project is to expand the knowledge on REV1 activity with a variety of lesions *in vitro*.

Methods and Algorithms: We expressed and purified the full-length human REV1 protein from a eukaryotic expression strain of *Saccharomyces cerevisiae* and characterized its DNA polymerase activity on undamaged DNA and a wide range of damaged DNA templates. The absence of impurities affecting the activity of REV1 was verified using the catalytically inactive REV1 variant. Human Y-family DNA polymerases Pol iota and Pol eta were used in control experiments to analyze Hoogsteen and Watson-Crick base interactions, respectively.

Results: G is known to be the preferable template of REV1 [3, 12–15]. We analyzed the effect of several biological and artificial modifications of template G. We showed that REV1 is blocked by the 1,2-GG cisplatin intrastrand crosslink, the predominant lesion caused by cisplatin chemotherapy. The artificial lesion constructed for Hoogsteen face impairment studies, 7-deazaG, also inhibited REV1, which is explained by the necessity of this face of the template nucleotide for contacts with the protein. Also, we analyzed REV1 activity on DNA with the most frequent lesion of G, 8-oxoG, with contradictory data in the literature [3, 13, 16, 17] and showed that REV1 is inhibited by over ~360 fold. At the same time, the result on the effect of O^6-meG inhibition (~36-fold inhibition) was in line with the literature data [3, 18].

A is the second preferred template of REV1 after G. By using Watson-Crick face impairing 1,N^6-ethenoA lesion and Hoogsteen face impairing 7-deazaA modification, we showed that REV1 is much more inhibited by Hoogsteen face impairment. We tested REV1 replication for 1,N^6-ethenoA with contradictory data in the literature [13, 19] and showed that REV1 is activated by over 50-fold, in accordance with [19]. Also, we studied 8-oxoA for the first time and showed that this oxidative lesion was readily replicated by REV1 in an error-prone manner.

Among the lesions of pyrimidines, while undamaged pyrimidines are the least favorite templates of REV1, methylcytosine (mC) and hydroxymethylcytosine (hmC) somewhat inhibited REV1. Surprisingly, thymine glycole (Tg) was replicated in an error-prone manner over 1700-fold more efficiently than undamaged T. It is important to note that most of the measured catalytic efficiencies (8-oxoG, O^6-meG, 1,N^6-ethenoA, Tg) showed that REV1 depends on discrimination at the binding step accompanied by a marked change in K_M and not k_{cat}.

Conclusion: In this work, we for the first time studied the activity of human REV1 on DNA containing the 1,2-GG cisplatin intrastrand crosslink, 7-deazaG and 7-deazaA nucleotide analogs, 8-oxoA and Tg lesions caused by oxidative stress as well as epigenetically modified nucleobases mC and hmC.

The structure of REV1 with templates A, T and C are still lacking in the literature. We provided biochemical evidence of the importance of the Hoogsteen face of template A for REV1 functioning. Template G is known to contact via its N7 and O^6 atoms of its Hoogsteen face with REV1 main chain amides His774 and Gly775 [4]. Our data suggest that the binding of templates G and A in the active site of REV1 is similar (i.e. A and G have similar conformations in the binding pocket).

REV1 was "activated" by 1,N^6-ethenoA and Tg lesions. Intact templates A, T and C are much poorer substrates for REV1 since they are unable to form interactions specific for the G base, where O^6 is a hydrogen bond acceptor. Increase in REV1 activity opposite 1,N^6-ethenoA can be explained by the loss of hydrogen bond donor properties of the atom in position 6 [19]. Moreover, based on our results and the literature data, addition of a CH3 group to C yielding mC as well as to U yielding T [13, 20, 21] decreases REV1 activity. However, the addition of OH directly to this group (to methylC yielding hydroxymethylC) or in its vicinity (to T yielding Tg) restores or even increases REV1 activity. Altogether, the obtained data shed light on the functioning of the active center of REV1 on different DNA substrates.

The results are partially available at https://www.mdpi.com/1422-0067/25/7/4107 (Stolyarenko, A., Novikova, A., Shilkin, E., Poltorachenko, V., Makarova, A., 2024. The Catalytic Activity of Human REV1 on Undamaged and Damaged DNA. International Journal of Molecular Sciences 25, 4107–4120. doi 10.3390/ijms25074107).

Funding: The study is supported by the state task of the National Research Center "Kurchatov Institute".

Список литературы/References

1. Bezalel-Buch R., Cheun Y.K., Roy U., Schärer O.D., Burgers P.M. Bypass of DNA interstrand crosslinks by a Rev1–DNA polymerase ζ complex. *Nucleic Acids Res*. 2020;48:8461-8473. doi 10.1093/nar/gkaa580
2. Ross A.-L., Simpson L.J., Sale J.E. Vertebrate DNA damage tolerance requires the C-terminus but not BRCT or transferase domains of REV1. *Nucleic Acids Res*. 2005;33:1280-1289. doi 10.1093/nar/gki279
3. Haracska L., Prakash S., Prakash L. Yeast Rev1 Protein Is a G Template-specific DNA Polymerase. *J Biol Chem*. 2002;277:15546-15551. doi 10.1074/jbc.M112146200
4. Swan M.K., Johnson R.E., Prakash L., Prakash S., Aggarwal A.K. Structure of the Human Rev1–DNA–dNTP Ternary Complex. *J Mol Biol*. 2009;390:699-709. doi 10.1016/j.jmb.2009.05.026
5. Shilkin E.S., Boldinova E.O., Stolyarenko A.D., Goncharova R.I., Chuprov-Netochin R.N., Khairullin R.F., Smal M.P., Makarova A.V. Translesion DNA Synthesis and Carcinogenesis. *Biochem Mosc*. 2020;85:425-435. doi 10.1134/S0006297920040033
6. Chan K., Resnick M.A., Gordenin D.A. The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. *DNA Repair*. 2013;12:878-889. doi 10.1016/j.dnarep.2013.07.008

7. Kim N., Mudrak S.V., Jinks-Robertson S. The dCMP transferase activity of yeast Rev1 is biologically relevant during the bypass of endogenously generated AP sites. *DNA Repair*. 2011;10:1262-1271. doi 10.1016/j.dnarep.2011.09.017
8. Otsuka C., Kunitomi N., Iwai S., Loakes D., Negishi K. Roles of the polymerase and BRCT domains of Rev1 protein in translesion DNA synthesis in yeast in vivo. *Mutat Res*. 2005;578:79-87. doi 10.1016/j.mrfmmm.2005.03.005
9. Zhou Y., Wang J., Zhang Y., Wang Z. The catalytic function of the Rev1 dCMP transferase is required in a lesion-specific manner for translesion synthesis and base damage-induced mutagenesis. *Nucleic Acids Res.* 2010;38:5036-5046. doi 10.1093/nar/gkq225
10. Sarkies P., Reams C., Simpson L.J., Sale J.E. Epigenetic Instability due to Defective Replication of Structured DNA. *Mol Cell*. 2010;40:703-713. doi 10.1016/j.molcel.2010.11.009
11. Jansen J.G., Langerak P., Tsaalbi-Shtylik A., Van Den Berk P., Jacobs H., De Wind N. Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *J Exp Med*. 2006;203:319-323. doi 10.1084/jem.20052227
12. Lin W., Xin H., Zhang Y., Wu X., Yuan F., Wang Z. The human REV1 gene codes for a DNA template-dependent dCMP transferase. *Nucleic Acids Res*. 1999;27:4468-4475. doi 10.1093/nar/27.22.4468
13. Zhang Y., Wu X., Rechkoblit O., Geacintov N.E., Taylor J.-S., Wang Z. Response of human REV1 to different DNA damage: preferential dCMP insertion opposite the lesion. *Nucleic Acids Res*. 2002;30:1630-1638. doi 10.1093/nar/30.7.1630
14. Masuda Y., Kamiya K. Biochemical properties of the human REV1 protein. *FEBS Lett*. 2002;520:88-92. doi 10.1016/S0014-5793(02)02773-4
15. Piao J., Masuda Y., Kamiya K. Specific amino acid residues are involved in substrate discrimination and template binding of human REV1 protein. *Biochem Biophys Res Commun*. 2010;392:140-144. doi 10.1016/j.bbrc.2009.12.167
16. Yeom M., Kim I.-H., Kim J.-K., Kang K., Eoff R.L., Guengerich F.P., Choi J.-Y. Effects of Twelve Germline Missense Variations on DNA Lesion and G-Quadruplex Bypass Activities of Human DNA Polymerase REV1. *Chem Res Toxicol*. 2016;29:367-379. doi 10.1021/acs.chemrestox.5b00513
17. Howel C.A., Prakash S., Washington M.T. Pre-Steady-State Kinetic Studies of Protein-Template-Directed Nucleotide Incorporation by the Yeast Rev1 Protein. *Biochemistry*. 2007;46:13451-13459. doi 10.1021/bi701429v
18. Choi J.-Y., Guengerich F.P. Kinetic Analysis of Translesion Synthesis Opposite Bulky N2- and O6-Alkylguanine DNA Adducts by Human DNA Polymerase REV1. *J Biol Chem*. 2008;283:23645-23655. doi 10.1074/jbc.M801686200
19. Yoon J.-H., Johnson R.E., Prakash L., Prakash S. DNA polymerase θ accomplishes translesion synthesis opposite 1,N 6 -ethenodeoxyadenosine with a remarkably high fidelity in human cells. *Genes Dev*. 2019;33:282-287. doi 10.1101/gad.320531.118
20. Masuda Y., Takahashi M., Fukuda S., Sumii M., Kamiya K. Mechanisms of dCMP Transferase Reactions Catalyzed by Mouse Rev1 Protein. *J Biol Chem*. 2002;277:3040-3046. doi 10.1074/jbc.M110149200
21. Acharya N., Haracska L., Prakash S., Prakash L. Complex Formation of Yeast Rev1 with DNA Polymerase η. *Mol Cell Biol*. 2007;27:8401-8408. doi 10.1128/MCB.01478-07

doi 10.18699/bgrs2024-11.1-10

PARP3 является потенциальным участником процесса регуляции структуры хроматина

Украинцев А.*, Кутузов М., Белоусова Е., Лаврик О.
Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** alan.ukraintsev@gmail.com*

Ключевые слова: нуклеосома; PARP1; PARP2; PARP3; АСМ

Мотивация и цель: В клетке ДНК находится в составе высокоорганизованного белок-нуклеинового комплекса – хроматина. Первым уровнем упаковки ДНК в составе хроматина является нуклеосома – комплекс, в котором ДНК обернута вокруг гистонов. ДНК-зависимые белки семейства ARTD – PARP1, PARP2, PARP3, принимают активное участие во многих клеточных процессах, связанных с хранением и передачей наследственной информации. Однако накоплено мало данных о природе взаимодействия белков PARP1/2/3 с нуклеосомой. Целью работы является изучить влияние этих трех белков на структуру нуклеосомы с помощью атомно-силовой микроскопии (АСМ).
Методы и алгоритмы: Модельная нуклеосома была реконструирована согласно [1] и состояла из ДНК длиной 346 п.н., содержащей нуклеосом-позиционирующую последовательность Clone 603 длиной 147 п. н. и два линкерных участка длиной 120 и 79 п. н., и октамерного комплекса гистонов – H2A, H2B, H3, H4.
В АСМ-исследованиях использовались слюдяные подложки. Для иммобилизации биологических образцов поверхность слюды модифицировали аминопропил силатраном. Сканирование модифицированной поверхности слюды с иммобилизованным препаратом проводили в режиме TappingMode на микроскопе Scanning Probe Microscope (Multimode 8, Veeco-Brucer, США). Обработку изображений и измерение геометрических параметров проводили с использованием программного обеспечения Gwyddion (v. 2.58) и ImageJ (v. 1.53e).
Результаты: Установлено, что белки PARP1/2/3 взаимодействуют с используемым в работе нуклеосомным субстратом. Оказалось, что все три белка взаимодействуют с модельной нуклеосомой в области входа-выхода линкерной ДНК из нуклеосомного диска. Обнаружено, что белок PARP3 уплотняет нуклеосому и при этом стабилизирует ее структуру. С белками PARP1/2 существенного влияния на компактизацию нуклеосомы обнаружено не было [2].
Выводы: Полученные результаты демонстрируют различные эффекты связывания белков PARP1/2/3 с модельным нуклеосомным субстратом. Новые данные о влиянии белка PARP3 на геометрию нуклеосомы позволяют предположить наличие функциональной активности в регуляции структуры хроматина этого белка.
Финансирование: Исследование поддержано РНФ (№ 22-74-10059).

PARP3 is a potential participant in the regulation of chromatin structure

Ukraintsev A.*, Kutuzov M., Belousova E., Lavrik O.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** alan.ukraintsev@gmail.com*

Key words: nucleosome; PARP1; PARP2; PARP3; AFM

Motivation and Aim: In a cell, DNA is contained in a highly organized protein-nucleic acid complex – chromatin. The first level of DNA packaging within chromatin is the nucleosome, a complex in which DNA is wrapped around histones. DNA-dependent proteins of the ARTD family – PARP1, PARP2, PARP3 – are actively involved in many cellular processes associated with the storage and transmission of hereditary information. However, little data has been accumulated on the nature of the interaction of the PARP1/2/3 proteins with the nucleosome. The goal of the work is to study the effect of these three proteins on the structure of the nucleosome using atomic force microscopy (AFM).
Methods and Algorithms: The model nucleosome was reconstitution according to [1] and consisted of DNA 346 bp long, containing the Clone 603 nucleosome positioning sequence 147 bp long and two linker regions 120 and 79 bp long, and an octamer histone complex – H2A, H2B, H3, H4.
Mica substrates were used in AFM studies. To immobilize biological samples, the mica surface was modified with aminopropyl silatrane. Scanning of the modified mica surface with the immobilized substance was carried out in TappingMode mode on a Scanning Probe Microscope (Multimode 8, Veeco-Brucer, USA). Image processing and measurement of geometric parameters were carried out using Gwyddion (v. 2.58) and ImageJ (v. 1.53e) software.
Results: It was established that PARP1/2/3 proteins interact with the nucleosomal substrate used in the work. It turned out that all three proteins interact with the model nucleosome in the region linker DNA entry-exit site of the nucleosome disk. Moreover, it was discovered that the PARP3 protein compacts the nucleosome and at the same time stabilizes its structure. No significant effect on nucleosome compaction was found with PARP1/2 proteins [2].
Conclusion: The results obtained demonstrate the different effects of binding of PARP1/2/3 proteins to a model nucleosomal substrate. New data on the influence of the PARP3 protein on nucleosome geometry suggest the presence of functional activity in the regulation of the chromatin structure of this protein.
Funding: The study is supported RSF (No. 22-74-10059).

Список литературы/References

1. Kutuzov M.M., Kurgina T.A., Belousova E.A., Khodyreva S.N., Lavrik O.I. Optimization of nucleosome assembly from histones and model DNAs and estimation of the reconstitution efficiency. *Biopolym Cell*. 2019;35:91-98
2. Ukraintsev A., Kutuzov M., Belousova E., Joyeau M., Golyshev V., Lomzov A., Lavrik O. PARP3 Affects Nucleosome Compaction Regulation. *Int J Mol Sci.* 2023;24(10):9042. doi 10.3390/ijms24109042

doi 10.18699/bgrs2024-11.1-11

Способность полимераз различных семейств процессировать субстраты, содержащие клик-химические модификации сахарофосфатного остова

Яковлев А.О.[1, 2]*, Ендуткин А.В.[1], Жарков Д.О.[1, 2]

[1] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *a.yakovlev1@g.nsu.ru*

Ключевые слова: клик-химия; ДНК-полимеразы; репликация ДНК; мутации

Мотивация и цель: «Клик-химия» – молодая и активно развивающая область химии, включающая в себя высокоэффективные, стерео- и региоселективные реакции, которые могут протекать при физиологических условиях. Наибольшее распространение клик-химии принесла реакция азид-алкинового циклоприсоединения, катализируемая ионами меди(I). Этот класс химических реакций широко используется для функционализации ДНК, РНК и соединений ненуклеотидной природы, модифицированных алкинильными группами. В этой связи стоит выделить крайне актуальное для молекулярной биологии клик-лигирование – соединение фрагментов нуклеиновых кислот с помощью 4-метил-1,2,3-триазольной группы (trz-группы). Межнуклеозидная trz-связь по своим свойствам радикально отличается от фосфатной группы – для применения конструктов, полученных клик-лигированием, необходимо исследовать особенности их процессинга белками, взаимодействующими с нуклеиновыми кислотами.

Для небольшого ряда полимераз было показано, что синтез ДНК по матрице, содержащей trz-связь, может проходить, при этом нуклеотидных замен обнаружено не было. Корректность синтеза была показана и в *E. coli* при трансформации плазмидной конструкцией, полученной клик-лигированием [1]. Следует отметить, что показанные результаты вступают в некоторое противоречие со структурными данными: ДНК в комплексе с ДНК-полимеразой в значительной степени изломана, что представляется более напряженной конформацией в случае межнуклеозидной триазольной группы [2]. В связи с этим исследование активности ДНК-полимераз в отношении продуктов клик-лигирования выступает актуальной задачей.

Методы и алгоритмы: Для проведения исследования были выбраны ДНК-полимеразы, относящиеся к трем семействам: A (фрагмент Кленова Pol I *E. coli* (KF)), B (RBpol и Pfu) и X (Pol β). В качестве удобной модели эукариотических полимераз (Pol δ и Pol ε) широкое распространение нашла полимераза фага RB69 ввиду высокой степени их гомологии.

Оценка способности выбранных полимераз синтезировать ДНК по матрице, содержащей trz-связь, была проведена с использованием двух типов конструкций для моделирования прохождения изучаемой модификации сахарофосфатного остова в двух режимах. В рамках первого исследовалась способность полимераз связывать субстрат с триазольной группой и инициировать синтез ДНК, второй же

режим репрезентировал элонгацию синтеза по модифицированной матрице. Кинетические параметры реакции включения нуклеотидов были определены в условиях стационарной кинетики.
Результаты: Полимеразы KF и RB69pol способны вести синтез в отношении содержащих trz-связь субстратов, однако было обнаружено, что KF в некоторой степени блокируется триазольной группой не только на стадии инициации, но и после включения первого нуклеотида. В свою очередь эффективность синтеза высокопроцессивной RB69pol оказалась значительно выше. Полимеразная активность Pfu, относящейся к тому же семейству, что и RB69pol, после включения первого за исследуемой модификацией нуклеотида полностью блокировалась. ДНК полимераза β достраивала праймер с очень низкой эффективностью, что также свидетельствует о подавлении синтеза ДНК trz-связью.
Было показано, что триазольная группа в составе сахарофосфатного остова стимулирует склонность полимераз к ошибкам. За trz-связью в использованной конструкции следует 5-метилцитозин, комплементарный dGMP, как и обычный цитозин. Однако все исследованные полимеразы, кроме Pol β, с сопоставимой эффективностью включали и другие нуклеотиды. Так, KF оказался способен встраивать в растущую цепь dAMP и dTMP, RB69pol, кроме корректного dGMP, инкорпорировала dAMP. Pfu, в свою очередь, включала только dAMP. Несмотря на низкоэффективный синтез, Pol β демонстрировала наиболее высокую точность среди выбранных полимераз и включала только корректный нуклеотид (dGMP).
Кинетические параметры реакций включения dAMP и dGMP полимеразами KF и RB69 также были определены. С использованием соотношений констант специфичности ферментативной реакции с корректным и некорректным нуклеотидом было показано, что RB69pol не только эффективнее проводит синтез по матрице, содержащей trz-связь, но и демонстрирует более высокую точность в сравнении с KF.
Выводы: Таким образом, возможность применения клик-лигирования в молекулярно-биологических целях вызывает определенные сомнения, так как использование этой технологии в сборке генов может привести к ошибкам полимераз, а в экспериментах *in vivo* провоцировать возникновение мутаций.
Финансирование: Исследование поддержано грантом Российского научного фонда № 21-74-10104.

The ability of polymerases of different families to process substrates containing click-chemical modifications of the sugar-phosphate backbone

Yakovlev A.O.[1, 2]*, Endutkin A.V.[1], Zharkov D.O.[1, 2]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *a.yakovlev1@g.nsu.ru*

Key words: click chemistry; DNA polymerases; replication; mutations

Motivation and Aim: “Click chemistry” is a young and rapidly developing field of chemistry that includes highly efficient, stereo- and regioselective reactions that can occur under physiological conditions. The most widespread reaction of click chemistry is the azide-alkyne cycloaddition catalyzed by copper(I) ions. This class of chemical reactions is widely used for the functionalization of DNA, RNA, and non-nucleotide compounds modified with alkyne groups. In this context, the connection of nucleic acid fragments using the 4-methyl-1,2,3-triazole group (trz-group), called click-ligation, is the highly relevant for molecular biology and should be highlighted. The inter-nucleoside trz-bond significantly differs in properties from the phosphate group – to apply constructs obtained by click ligation it is necessary to investigate the peculiarities of their processing by proteins interacting with nucleic acids.

For a small set of polymerases, it has been shown that DNA synthesis on a template containing a trz-bond can proceed without nucleotide substitutions. The correctness of synthesis was demonstrated in E. coli during transformation with a plasmid construct obtained by click-ligation [1]. It should be noted that these results somewhat contradict structural data: DNA in complex with DNA polymerase is significantly bent, which seems to be a more strained conformation in the case of the inter-nucleoside triazole group [2]. Therefore, studying the activity of DNA polymerases towards click ligation products emerges as a relevant task.

Methods and Algorithms: For the study, DNA polymerases belonging to three families were selected: A (Klenow fragment of *E. coli* Pol I (KF)), B (RBpol and Pfu) and X (Pol β). As a convenient model of eukaryotic polymerases (Pol δ and Pol ε), the polymerase of phage RB69 has gained wide popularity due to high degree of their homology.

The ability of the selected polymerases to synthesize DNA on a template containing a trz-bond was evaluated using two types of constructs to simulate the passage of the studied sugar-phosphate backbone modification in two modes. The first mode examined the polymerases' ability to bind substrate with a triazole group and initiate DNA synthesis, while the second mode represented elongation of synthesis on the modified template. Kinetic parameters of nucleotide incorporation reaction were determined under conditions of steady-state kinetics.

Results: Polymerases KF and RB69pol are capable of synthesizing substrates containing trz-bonds, but it has been observed that KF is partially blocked by the triazole group not only during initiation but also after the inclusion of the first nucleotide. In contrast, the synthesis efficiency of the highly processive RB69pol was significantly higher. The polymerase activity of Pfu, belonging to the same family as RB69pol, was completely blocked after the inclusion of the first modified nucleotide studied. DNA polymerase β extended the primer with very low efficiency, indicating suppression of DNA synthesis by the trz-bond.

It has been shown that the triazole group in the sugar-phosphate backbone stimulates polymerases to make errors. Following the trz-bond in the used construct there is 5-methylcytosine complementary to dGMP, similar to the usual cytosine. However, all polymerases examined, except for Pol β, which showed comparable efficiency, also included other nucleotides. For example, KF was able to incorporate dAMP and dTMP into the growing chain, RB69pol incorporated dAMP in addition to the correct dGMP. Pfu, on the other hand, only included dAMP. Despite low-efficiency synthesis, Pol β demonstrated the highest accuracy among the selected polymerases and included only the correct nucleotide (dGMP).

Kinetic parameters of inclusion reactions of dAMP and dGMP by KF and RB69 polymerases were also determined. Using specificity constants for the enzymatic reaction with the correct and incorrect nucleotide, it was shown that RB69pol not only conducts synthesis more efficiently on a template containing trz-bonds but also demonstrates higher accuracy compared to KF.

Conclusion: Thus, the feasibility of using click ligation for molecular biological purposes raises some doubts, as the use of this technology in gene assembly may lead to polymerase errors and, in *in vivo* experiments, provoke the occurrence of mutations.

Funding: The study is supported by Russian Science Foundation (grant No. 21-74-10104).

Список литературы/References

1. El-Sagheer A.H. et al. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli. *Proc Natl Acad Sci USA*. 2011;108(28):11338-11343
2. Rothwell P.J., Waksman G. Structure and mechanism of DNA polymerases. *Adv Protein Chem*. 2005;71:401-440

doi 10.18699/bgrs2024-11.1-12

Analysis of critical amino acid residues in the active center of a novel DNA glycosylase Fpg-like protein 1 *Streptomyces coelicolor* superfamily "helix-2-turn-helix" DNA glycosylase

Bulgakov N.A.[1, 2]*, YudkinaA.V.[1, 2, 3], Zharkov D.O.[1, 2], Garcia-Diaz M.[3]

[1] *SB RAS Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Stony Brook University, Stony Brook, USA*

* *n.bulgakov@alumni.nsu.ru*

Key words: DNA damage; DNA repair; formamidopyrimidine–DNA glycosylase; endonuclease VIII; protein crystalology; *Streptomyces coelicolor*; protein crystallography

Motivation and Aim: There is DNA in living organisms under the constant action of genotoxic agents. These agents could be internalized in cell (radiation or UV radiation) or appear like cell's reactive metabolites (for example reactive oxygen species). To maintain the integrity of the genome, cells have repair systems. One of these systems is the base excision repair system, which is responsible for removing the most common DNA damage, oxidative damage to nucleobases. The major DNA glycosylase of the bacterial cell is formamidopyrimidine-DNA glycosylase (Fpg), which recognizes and removes 8-oxoguanine from DNA, a marker of oxidative stress. Today, DNA–glycosylases are divided into several superfamilies on the basis of the enzymes structural homology. Fpg belongs to the helix-2-turn-helix superfamily of DNA–glycosylases. Another key member of this superfamily DNA-glycosylase Nei – recognizes and removes oxidized pyrimidines.

Recent full genome sequencing of a number of bacteria has shown that there are bacteria with 4 homologs of DNA–glycosylases of the «helix-2-turn-helix» superfamily in their genome, whereas normally there are 1-2 glycosylases of this superfamily in the bacterial cell. Alignment of the aminoacid residues of the new Fpg DNA–glycosylases showed altered aminoacid residues in both the active center of the enzyme and the recognition center of the damaged base. Such mutations may affect not only the substrate specificity of the enzyme but also the quality of the enzyme-DNA interaction. The study of Fpg-like proteins will allow us to understand the contribution of amino acids not only in catalysis but also in the recognition of damaged bases, which may allow us to create a tool for designing enzymes with a given substrate specificity.

The aim of this work is to biochemically characterize the enzyme Fpg-like protein 1 *Streptomyces coelicolor*, and to establish the spatial structure of the free enzyme using protein crystallography to further characterize the mechanism of the enzyme using physicochemical methods.

Methods and Algorithms: We have used molecular cloning and optimized protocol of protein purification to get new Fpg-like proteins. Using biochemical approaches, we have got first understanding of affinity and substrate specificity of these enzymes. By protein crystallography method we have crystals and resolved the structures of new Fpg-like proteins from *Streptomyces coelicolor* and *Bacteroides thetaiotaomicron*.

Results: In the molecule of «helix-2-turn-helix» superfamily belonging proteins there are some structural features that provide the mechanism of recognize, affinity and catalysis. Most “helix-2-turn-helix” glycosylases have high-conservative PELPEVET N-terminal motif and use their N-terminal proline residue as a key catalytic nucleophile. Additionally, «helix-2-turn-helix» proteins have helix-two-turn-helix domain and zinc-finger motif to contribute residue to the active site. However, new Fpg-like proteins have significant diversity of N-terminal motif and residue recognizing damage. We have crystallized Fpg-like protein from *Streptomyces coelicolor* with 1.63 Å resolution and reveal the presence of Zn-finger domain in C-terminal of protein, α-helix rich region in N-terminal and some another Fpg-like features. However, substrate preferences make these proteins similar to Nei.

Conclusion: We have revealed structural and biochemical aspects of new atypical Fpg-like proteins to determine relationships these proteins to known glycosylases families by protein crystallographic approach and also we have investigated aminoacid substitutions in Fpg *E. coli* to estimate contribution distinguished active center aminoacids in catalysis, affinity and recognition Fpg-like proteins with respect to Fpg wild type.

Funding: The study is supported R&D contract No. 864/22/N between Federal Research Center Boreskov Institute of Catalysis SB RAS and Institute of Chemical Biology and Fundamental Medicine SB RAS

The authors thank Dr. M. Saparbaev for plasmids containing new Fpg-like proteins. Also The authors thank Dr. V. Borshevsky and M. Shevtsov for their help in obtaining Fpg-like protein 1 crystals, collecting diffraction data and further interpretation of these data.

doi 10.18699/bgrs2024-11.1-13

Repair efficiency of biorthogonal DNA modifications in human cells

Kim D.V.*, Endutkin A.V., Zharkov T.D., Zharkov D.O.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** kim.daria.nsk@gmail.com*

Key words: biorthogonal; repair; HeLa; reporter system

Motivation and Aim: Non-natural nucleotides are efficiently used in biotechnology and biomedicine studies to increase the stability of DNA oligonucleotides within the cell or organism. Nevertheless, little is known about the interaction of cellular systems with DNA containing non-natural nucleotides. This study aimed to investigate DNA repair of non-natural nucleotides in human cells.
Methods and Algorithms: The repair of biorthogonal DNA modifications was studied using the reporter system. This reporter system was based on the plasmid, containing the sequence of non-fluorescent variant of EGFP [1]. The modification was introduced in the transcribed strand of the *EGFP* sequence. The repair of modification within the cell would restore the original *EGFP* sequence if not RNA polymerase II would incorporate random ribonucleotide opposite to the lesion leading to EGFP fluorescence. To sum up, the level of EGFP fluorescence was in inverse ratio to DNA repair.
Results: The repair of several non-natural DNA modifications were studied such as the enantiomers of natural nucleotides (βLdN), alkylated derivatives of uracil (5etU, C8AlkU) and alkylated abasic site (1etF). The plasmid containing the modification was transfected with a plasmid containing red fluorescent protein as a transfection efficiency control in HeLa cells. Cells were analyzed 24 h after transfection. It was shown that the DNA repair efficiency of βLdN varies. The pyrimidine L-enantiomers were repaired less efficiently than purine L-nucleotides. The 5etU and C8AlkU predominantly did not undergo repair in human cells. 1etF was repaired as efficiently as a synthetic analogue of an abasic site. 1etF was conjugated with a non-bulky heterocyclic ligand that resembles a DNA base. This click-conjugation of triazine derivative with 1etF was more resistant to repair than 1etF.
Conclusion: Non-natural DNA modifications are important tools for basic research and nanotechnology. Thus, it is necessary to investigate the interaction of the cellular machinery with these modifications. We demonstrated that several non-natural DNA modifications were repaired within the cell.
Funding: The research was supported by the Russian Science Foundation (21-74-10104).

References

1. Kitsera N., Rodriguez-Alvarez M., Emmert S., Carell T., Khobta A. Nucleotide excision repair of abasic DNA lesions. *Nucleic Acids Res*. 2019;47(16):8537-8547

doi 10.18699/bgrs2024-11.1-14

The role of Drosophila Rad51D in double strand breaks repair

Konev A.Y.[1]*, Il'ina Yu.A.[1], Ukraintsev V.Y.[1], Schevlyakov A.D.[1],
Toroshchina A.V.[1], Kozyreva S.Y.[2], Fishman V.S.[2]

[1] *Petersburg Nuclear Physics Institute named by B.P. Konstantinov of National Research Centre "Kurchatov Institute", Gatchina, Russia*

[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

* *konev_ay@pnpi.nrcki.ru*

Key words: DNA repair; homologous recombination; *Drosophila*; Rad51D, Rad51 paralogs

Motivation and Aim: Cells use several mechanisms to repair DNA double – strand breaks (DSB): homologous recombination (HR), single-strand annealing (SSA), canonical and alternative pathways of non-homologous end joining (NHEJ). The key event in recombination repair of DNA double strand breaks (DSB) is the formation of the Rad51 nucleoprotein filament, which is necessary for the homology search and exchange of DNA strands. An important role in the regulation of the assembly, stabilization, and disassembly of Rad51 filaments is played by mediator proteins, including the Rad51 paralogs – proteins structurally similar to Rad51. Mammals have six paralogs of the Rad51: Rad51B, Rad51C, Rad51D, XRCC2, XRCC3 and the SWSAP1 (RadA homolog). Germline mutations in RAD51 paralogs cause predisposition to breast and ovarian cancers. Despite their significant role in DSB repair, the individual functions of Rad51 paralogs are not fully understood. Drosophila is a classical multicellular eukaryotic model for studies of DNA damage repair at the organismal level, but our knowledge of the functions of Rad51 paralogs in this organism is lagging behind. Drosophila genome contains the Rad51 ortholog *spnA* and four Rad51 paralogs: *spnB* (XRCC3 homolog), *spnD* (hRad51C), *XRCC2* and *Rad51D*. The $rad201^{G1}$ mutation (radiation sensitive 201) was isolated from a natural population by its larval hypersensitivity to ionizing radiation [1]. Here we show that the $rad201^{G1}$ mutation is a *Rad51D* allele and use it to study of the role of the Drosophila *Rad51D* in DSB repair.
Methods and Algorithms: The *Rad51D* expression level was determined by qPCR; actin and rosy genes were used as references. To analyze the role of the *Rad51D* in DSB repair we employed a single DSB system utilizing DR-*white* (direct repeat *white*) reporter [2]. The DR-white transgene has one upstream nonfunctional *white* gene containing an I-SceI site, a red fluorescent marker, and a downstream truncated non-functional white gene (*iwhite*). To produce DSB we used I-SceI fused to an ecDHFR degradation domain which allows I-SceI protein stabilization after addition of the ligand trimethoprim to fly food. To visualize DSBs in the living tissue, these flies also express mu2 tagged with enhanced yellow fluorescent which binds to γH2Av and is recruited to DSB. The kinetics of mu2-eYFP foci disappearance after irradiation was analyzed in an isolated cultivated brain of a late Drosophila larva under Leica scanning microscope. In order to analyze the spectrum of repair products in somatic cells we induced breaks in flies containing DR-*white* and I-SceI transgenes and subsequently PCR-amplified the upstream white gene from whole larval genomic DNA and identified the exact sequences present in all products using MGI sequencing of DR-*white* PCR products. Reads alignment were

performed using bwa. Subsequent data analysis was performed using custom python script based on CIGAR information provided in sam-files.

Results: Here we show that the *rad201*^{G1} mutation is caused by the insertion of the Opus retrotransposon at the 5' untranslated region of the *Rad51D* gene. In addition to the Opus insertion in the site of mutation, the "*rad201*^{G1}" chromosome contains a number of nucleotide changes, which cause K61E, V93A and Y108H amino acid substitutions in the Rad51D and F50L in the protein encoded by the overlapping *CG42382* gene. We isolated spontaneous reversions of the *rad201* radiation sensitivity phenotype. All reversions are associated with the loss of Opus, leaving the nucleotide substitutions in Rad51D and CG42382 genes intact. In the *rad201*^{G1} mutant embryos the *Rad51D* transcription is 30-fold reduced by contrast with the wild type and is 20-fold reduced in 3-day virgin female ovaries while the *CG42382* transcription does not change, supporting that *rad201*^{G1} is the Rad51D allele. We have examined the *Rad51D* expression level in 4–6 hour Drosophila embryos 30 min and 120 min after exposure to 7 Gy of γ-rays and in 3-day virgin female ovaries after 30 min and 120 min after exposure to 20 Gy. Both in wild type embryos and in *rad201*^{G1} mutants the *Rad51D* expression level remains unchanged 30 min after exposure to 7 Gy γ-rays. After 120 min the *Rad51D* expression increases 2-3 times in wild embryo and increases 2 times in 3-day virgin female ovaries. These data indicate the up-regulation of the *Rad51D* gene expression level in response to gamma-ray irradiation.

The *rad201*^{G1} mutants are sensitive to ionizing irradiation and to cisplatin, but not to MMS or UV-light. By contrast with the other studied members of the Rad51 family in Drosophila, *Rad51D* mutant has a rather weak spindle phenotype which appears only with age, suggesting some limited role in meiotic DSB repair. A live analysis of the kinetics of radiation-induced DNA damages repair in the neuroblasts nuclei of isolated alive larval brains was performed by monitoring of mu2–yYFP foci disappearance with time. We have shown that the kinetics of DNA repair is delayed in drosophila *Rad51D* mutant.

Repair of the single Sce-I induced DSB by HR, NHEJ, and SSA is quantitated by determining the phenotypes of the progeny (male germline repair events) or by PCR and sequence analysis in somatic cells. In order directly determine which repair pathways affected in Drosophila *Rad51D* mutant we employed the DR-*white* system to quantitate the frequencies of different DSB repair products in the male germline. Single DSBs were induced in both premeiotic germ cells and somatic cells by feeding DR-white/ I-SceI-ecDHFR larvae with trimethoprim [2]. Red-eyed DR-white progeny indicate an HR repair event in the paternal male germline, white eyed flies with loss of the intervening 3xP3.dsRed marker results from SSA and flies with white, but fluorescent eyes result from the absence of an I-SceI-induced DSB, NHEJ or HR repair using the sister chromatid. Our results show that the frequency of HR is 4 times reduced in Drosophila *Rad51* mutants germline with corresponding increase in SSA events by contrast with wild type animals.

Next, we determined the proportion of HR and NHEJ events in somatic cells and identified the exact sequences present in all DR-*white* PCR products using NGS sequencing. Result of HR in this case looks as 23 bp deficiency of inserted into upstream white gene sequence. Mutation *rad201*^{G1} of *Rad51D* gene reduces the frequency of HR from 21 to 4 % in somatic cells comparing to wild type flies and increased proportion of indels. Thus, like Rad51 (*spnA*) Drosophila Rad51D is required for DSB repair by HR mechanism in Drosophila both in germline and somatic cells. However by contrast with

Rad51 mutations or depletion, where reduction of HR caused general decrease of repair products [2, 3] *rad201*^{G1} mutation leads to increased production of indels resulted from NHEJ.

Conclusion: We identified *rad201*^{G1} mutation as an allele of Drosophila *Rad51D* gene. Rad51D is required for DSB by mechanism of homologous recombination both in germline and somatic cells. *rad201*^{G1} mutant phenotype is similar to the phenotype of Drosophila *Xrcc2* mutant, suggesting that Rad51D – Xrcc2 proteins might form a complex required for DSB repair in somatic cells.

Funding: The research was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation (theme №1023031500033-1-1.6.7;1.6.4;1.6.8).

References

1. Khromykh I.M., Zakharow I.A. Studies on *Drosophila* radiosensitive strains. Identification of the second chromosome-linked mutant genes controlling somatic mutagen sensitivity. *Genetika*. 1981;17(4):658-666 (in Russian)
2. Janssen A., Breuer G.A., Brinkman E.K., van der Meulen A.I., Borden S.V., van Steensel B., Bindra R.S., LaRocque J.R., Karpen G.H. A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev.* 2016;30(14):1645-1657. doi 10.1101/gad.283028.116
3. Do A.T., Brooks J.T., Le Neveu M.K., LaRocque J.R. Double-strand break repair assays determine pathway choice and structure of gene conversion events in *Drosophila melanogaster*. *G3 (Bethesda).* 2014;4(3):425-432. doi 10.1534/g3.113.010074

doi 10.18699/bgrs2024-11.1-15

Intra-strand DNA symmetry provides insight into evolution of complex life form

Saparbaev M.

Group "Mechanisms of DNA Repair and Carcinogenesis", CNRS UMR9019, Université Paris-Saclay, Gustave Roussy Cancer Campus, F-94805 Villejuif Cedex, France

A major challenge of modern biology is to understand how random genetic variations cause phenotypic variation in a quantitative trait. Francis Crick's central dogma provides residue-by-residue mechanistic explanation of the flow of genetic information in living systems. However, this principle may not be sufficient to address the above problem. Chargaff's Second Parity Rule (CSPR), also referred to as intra-strand DNA symmetry, and defined, as near exact equalities G≈C and A≈T within single DNA strand, is a statistical property of cellular genomes. Analysis of mutation spectra inferred from single nucleotide polymorphisms observed in human and mice populations revealed near exact equalities of the reverse complementary mutations, indicating that random genetic variations are in compliance with CSPR. Furthermore, the nucleotide compositions of coding sequences are statistically interwoven via CSPR, since pyrimidine bias at 3rd codon position compensates purine bias at 1st and 2nd positions. Based on Fisher's infinitesimal model, we propose that accumulation of reverse complementary mutations results in the continuous phenotypic variation due to small additive effects of statistically interwoven genetic variations. Therefore, additive genetic interactions can be inferred as a statistical entanglement of the nucleotide compositions of separate genetic loci. CSPR challenges neutral theory of molecular evolution, since all random mutations participate in variation of a trait and provides alternative solution to Haldane's dilemma by making a gene function diffuse. We propose that CSPR is a symmetry of Fisher's infinitesimal model and that genetic information can be transferred in implicit, contactless manner. Natural selection acts on many traits at once, consequently, evolution proceeds via accumulation of small effect-size mutations distributed over entire genome. Statistical entanglement of random genetic variations is the way to store information despite genetic drift.

doi 10.18699/bgrs2024-11.1-16

G-quadruplexes at super-enhancers and CTCF clusters help shape topologically associating domains/replication domains

Varizhuk A.*, Pavlova Iu., Iudin M., Barinov N., Sultanov R., Bogomazova A., Klinov D., Tsvetkov V.

Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine, Moscow, Russia
** annavarizhuk@gmail.com*

Key words: G-quadruplex; topologically associating domain; replication domain; CTCF; nucleosome density; super-enhancer; BRD4

Motivation and Aim: Topologically associating domains (TADs), which are established in the G1 phase of the eukaryotic cell cycle and restrict the enhancer-promoter communication, coincide with replication domains, which determine replication timing in the S-phase. Several models of the spatiotemporal regulation of transcription and replication have been proposed to explain this observation. These include the Rosette model and the CRISI (chromatin reorganization induced selective initiation) model [1]. The former suggests the intra-TAD co-firing of high-efficiency replication origins, whereas the latter implies the relocation of the high-efficiency origins towards the TAD periphery due to TAD dynamics upon the enhancer-promoter communication. Both models provide new insights into the regulatory role of G-quadruplex DNA structures (G4s) in replication. So far, they have only been considered as possible sites of pre-replication complex assembly or replication stress triggers in helicase-deficient cells. However, a growing body of data suggests that G4s may also be involved in replication-associated chromatin remodeling, namely TAD demarcation and dynamics. The former process requires local accumulation of the CCCT-binding factor (CTCF), and the latter can be driven by liquid-liqiud phase separation (LLPS) at super-enhancers (SEs). In light of the reported colocalization of G4s with strong TAD boundaries and intra-TAD interaction hotspots in the human genome [2], we sought to investigate whether G4s promote the recruitment of CTCF or LLPS drivers.

Methods and Results: Recently, we investigated the impact of G4s on CTCF positioning based on CTCF-ChIP-seq experiments using human myeloid leukemia cells (K562) pretreated with a G4-stabilizing ligand, pyridostatin (PDS). The CTCF profiles in control or PDS-treated cells were compared to the reported profiles of G4s predicted by G4-seq or confirmed by G4-ChIP-seq. The G4s with confirmed intrinsic stability in the K562 context demonstrated a high correlation with CTCF in the control samples, with minimal response to the stabilizing ligand. Conversely, the predicted (semi-stable) G4s exhibited a weak correlation with CTCF in the control samples, but this correlation increased significantly in response to PDS [3]. The ChIP-seq results were corroborated by ChIP-PCR for several representative G4 sites, and some of the G4s were shown to bind CTCF with Kd values comparable to that obtained for the consensus CTCF binding motif in a microscale thermophoresis (MST)-based binding assay.

These findings agree with the results obtained by other groups [4] and support direct CTCF recruitment by G4s. Importantly, they do not rule out indirect G4-related mechanisms of CTCF accumulation. In particular, G4s could facilitate the positioning of CTCF at nearby cognate binding sites by modulating the local CpG methylation rates

and/or nucleosome density. The molecular basis for G4-dependent methylation control has been explored elsewhere. We aimed to elucidate the relationship between G4s and nucleosome-depleted regions (NDR).

We obtained a minimal chromatin model, which included a single histone octamer (HO) assembled on a dsDNA comprising a core Widom 601 nucleosome positioning sequence and a G4-harboring or non-G4 tail. The folding of the G4 within the dsDNA tail was confirmed by several methods, including thioflavin light-up assays and ^{1}H-NMR. HO assembly on dsDNA with G4 versus non-G4 tails was analyzed by polyacrylamide gel electrophoresis. The relative intensities of the nucleosome bands indicated a reduced efficiency of HO assembly near the G4 tail. This observation was additionally verified by atomic force microscopy (AFM)-based analysis of nucleosome morphologies in samples with G4 versus non-G4 DNA [5]. Taken together, our data confirm nucleosome exclusion around G4s, which may be favorable for CTCF clustering at TAD boundaries but would also affect the intra-TAD connectivity. Given that transcriptionally active TADs are typically hyperacetylated (H3K27ac) at G4-rich promoter and enhancer regions, we additionally verified the exclusion of HOs with acetylated H3 around G4s and obtained analogous results.

In studying intra-TAD interactions, we focused on the formation biomolecular condensates at SEs, which typically communicate with multiple promoters. Such biocondensates are appear to be scaffolded by the SE marker BRD4, which recognizes acetylated chromatin through its tandem bromodomains and co-separates with the Mediator complex subunits due to its disordered C-terminal region [6]. The analysis of BRD4 distribution in K562 SEs using ChIP-seq data revealed high occupancy of this protein at/near NDRs, which are twice as frequent in G4-rich SEs as in G4-free SEs. Since acetylated nucleosomes are absent around G4s, the mechanism of BRD4 maintenance is currently uncertain. LLPS explains the nearly constant concentration of BRD4 along the SE. However, LLPS necessitates that DNA engages in multiple transient (weak) interactions with BRD4. Nucleosome-free dsDNA does not bind to BRD4 or facilitate LLPS at biologically relevant concentrations. Therefore, we tested representative G4s from SEs for BRD4 binding and LLPS promotion.

Representative G4s from K562 SEs were identified using MEME and FIMO algorithms. Sequences matching the highly enriched and moderately enriched motifs were selected for binding assays. Negative controls included dsDNA and iM DNA. Both G4s showed weak affinity for BRD4 with med-micromolar Kd values in both MST-based and surface plasmon resonance (SPR)-based assays, while dsDNA and iM showed negligible binding. The highly enriched motif (hereafter referred to as SE-G4) bound BRD4 more efficiently than the moderately enriched motif in both MST and SPR assays.

To assess the impact of SE-G4 on BRD4-driven LLPS, fluorescently labeled BRD4 was mixed with isolated SE-G4 or SE-G4 within dsDNA (nucleosome-free), DNA-free acetylated histones (negative control), nucleosome-free non-G4 dsDNA (negative control), or acetylated nucleosomes assembled on non-G4 dsDNA (positive control). LLPS was monitored by fluorescence microscopy. Biocondensates, visualized as micrometer-scale droplets, were observed only in the positive control and G4-containing samples. Their morphology was confirmed by AFM. To summarize, G4s from SEs facilitated LLPS *in vitro* and thus may contribute to the scaffolding of biocondensates at SEs.

Conclusion: We have shown that G4 structures, which are enriched at TAD boundaries or intra-TAD interaction hotspots, can directly recruit CTCF or increase the accessibility

of its binding sites within dsDNA by excluding nearby nucleosomes. At G4-rich SEs within transcriptionally active TADs, the exclusion of acetylated nucleosomes critical for BRD4 loading is compensated by the increased LLPS-promoting potency of G4s compared to duplex DNA. Our results add to the current understanding of G4 involvement in TAD demarcation and enhancer-promoter communication. However, it should be noted that they were obtained using simplified chromatin models and require further verification.

Funding: The study is supported by Russian Science Foundation (No. 22-15-00129).

References

1. Li Y., Xue B., Zhang M. et al. Transcription-coupled structural dynamics of topologically associating domains regulate replication origin efficiency. *Genome Biol*. 2021;22(1):206. doi 10.1186/s13059-021-02424-w
2. Hou Y., Li F., Zhang R., Li S., Liu H., Qin Z.S., Sun X. Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. *Epigenetics*. 2019;14(9):894-911. doi 10.1080/15592294.2019.1621140
3. Tikhonova P., Pavlova Iu., Isaakova E. et al. DNA G-Quadruplexes Contribute to CTCF Recruitment. *Int J Mol Sci*. 2021;22(13):7090. doi 10.3390/ijms22137090
4. Roy S.S., Bagri S., Sengupta A., Then C.R., Kumar R., Sridharan S., Chowdhury S. Artificially inserted G-quadruplex DNA secondary structures induce long-distance chromatin activation. *eLife*. 2024;13:RP96216. doi 10.7554/eLife.96216.1
5. Pavlova I., Barinov N., Novikov R. et al. Modeling G4s in chromatin context confirms partial nucleosome exclusion and reveals nucleosome-disrupting effects of the least selective G4 ligands. *Biochimie*. 2023;204:8-21. doi 10.1016/j.biochi.2022.08.016
6. Sabari B.R., Dall'Agnese A., Boija A. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. 2018;361(6400):eaar3958. doi 10.1126/science.aar3958

doi 10.18699/bgrs2024-11.1-17

Order and disorder in base excision DNA repair

Zharkov D.O.[1, 2]*

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

* *dzharkov@niboch.nsc.ru*

Key words: DNA repair; DNA glycosylases; AP endonucleases; intrinsically disordered regions; protein dynamics; substrate recognition; multifunctional proteins

Motivation and Aim: Structures of many polypeptides possess intrinsically disordered regions, which may be important for protein–protein and enzyme–substrate interactions, protein intracellular localization, etc. [1, 2]. A number of such proteins may be found in base excision repair (BER), a pathway that purges our genome of most lesions generated in DNA by spontaneous and induced oxidation, deamination, alkylation, and base loss [3, 4]. However, the roles of the disordered regions in these proteins are mostly unknown.

Methods and Algorithms: We have used X-ray crystallography and computational methods (molecular dynamics, coarse-grained Monte Carlo simulations, *ab initio* folding, molecular docking, etc.) to determine experimentally or predict structures of several human and bacterial DNA glycosylases and apurinic/apyrimidinic (AP) endonucleases, the enzymes that catalyze the first two steps of the BER pathway. Direct enzyme kinetics, binding and substrate perturbation methods combined with site-directed mutagenesis were used to verify functional predictions made based on the structural information.

Results: Three main structural classes of DNA glycosylases are α/β-fold uracil–DNA glycosylase-like (UNG-like) proteins (*Escherichia coli* Ung and Mug, human UNG, TDG, and SMUG1), helix–hairpin–helix (HhH) motif containing proteins (*E. coli* Nth, MutY, and AlkA, human OGG1, NTHL1, MUTYH, and MBD4), and helix–two-turn–helix (H2TH) motif containing proteins (*E. coli* Fpg and Nei, human NEIL1, NEIL2, and NEIL3) [4, 5]. AP endonucleases belong to exonuclease–endonuclease–phosphatase (EEP) structural family (*E. coli* Xth, human APEX1) and triose isomerase barrel (TIM barrel) structural family (*E. coli* Nfo) [4, 5]. Most human homologs have extended N- and C-terminal tails while their bacterial counterparts are limited to the catalytic core. Even in two-domain bacterial H2TH glycosylases Fpg and Nei, the linker between the domains is well-ordered and restructures itself for substrate DNA binding. Surprizingly, the most disordered region in Fpg and Nei was their substrate-binding pocket, which may explain profound substrate promiscuity of these enzymes in terms of the lesions recognized. The terminal tails in human DNA glycosylases and APEX1 are predicted as disordered and are not seen in the X-ray structures but molecular modeling reveals their tendency to exist in several conformations rather than be completely disordered. Some of these conformations are well suitable for facilitating DNA binding or protein multimerization on DNA. Several human DNA glycosylases (NEIL2, MBD4, MUTYH) possess long internal unstructured regions, which we found to possess autoinhibitory properties possibly relieved upon interaction with other BER participants or with specific DNA elements such as methylated CpG sites.

Conclusion: Disordered regions in DNA glycosylases and AP endonucleases are involved in non-specific DNA binding, discrimination of damaged *vs* undamaged bases, and regulation of enzyme activity. Unlike truly intrinsically disordered regions, many of them likely populate a limited number of distinct conformations, which could have specific functional roles.

Funding: The study is supported by Russian Science Foundation (No. 24-14-00285) and the Strategic Academic Leadership Priority 2030 Program of Novosibirsk State University.

References

1. Oldfield C.J., Dunker A.K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem.* 2014;83:553-584. doi 10.1146/annurev-biochem-072711-164947
2. van der Lee R., Buljan M., Lang B. et al. Classification of intrinsically disordered regions and proteins. *Chem Rev.* 2014;114(13):6589-6631. doi 10.1021/cr400525m
3. David S.S., O'Shea V.L., Kundu S. Base-excision repair of oxidative DNA damage. *Nature.* 2007;447(7147):941-950. doi 10.1038/nature05978
4. Zharkov D.O. Base excision DNA repair. *Cell Mol Life Sci.* 2008;65(10):1544-1565. doi 10.1007/s00018-008-7543-2
5. Hitomi K., Iwai S., Tainer J.A. The intricate structural chemistry of base excision repair machinery: Implications for DNA damage recognition, removal, and repair. *DNA Repair.* 2007;6(4):410-428. doi 10.1016/j.dnarep.2006.10.004

11

Симпозиум «Фундаментальные генетические/клеточные системы/процессы: компьютерные и экспериментальные подходы»

Symposium “Fundamental genetic/cellular systems/processes: computational and experimental approaches”

doi 10.18699/bgrs2024-11.2-01

Регуляция терминации трансляции эукариот факторами инициаторного комплекса

Алкалаева Е.З.*, Шувалова Е.Ю., Шувалов А.В., Бизяев Н.С., Аль Шейх В.
Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия
** alkalaeva@eimb.ru*

Ключевые слова: терминация трансляции; инициация трансляции; рибосома

Мотивация и цель: Эукариотический терминационный комплекс помимо основных двух факторов терминации eRF1 и eRF3 может включать в себя гораздо больше белков. Так, у дрожжей и млекопитающих была показана регуляция терминации трансляции белками PABP, eIF3j, DDX19, Gle1, eIF5A, ABCE1, PDCD4, PAIP1 и PAIP2. Кроме того, очевидно, что в ряде процессов, протекающих во время трансляции, таких как образование замкнутой closed-loop структуры мРНК и трансляция на коротких uORFs с последующей реинициацией трансляции, инициаторные и терминационные рибосомные комплексы сближаются. Таким образом, факторы инициации потенциально могут влиять на активность факторов терминации трансляции. Целью данной работы было комплексное исследование роли факторов инициации в терминации трансляции эукариот.
Методы и алгоритмы: Для решения поставленной задачи нами была использованы модельные *in vitro* системы: реконструированная из отдельных компонентов система трансляции млекопитающих и очищенные из лизата ретикулоцитов кролика претерминационные комплексы. С помощью данных методов было исследовано влияние факторов инициации трансляции на разные стадии терминации трансляции.
Результаты: Нами обнаружено, что фактор инициации eIF4F, связывающийся с PABP, а точнее две его субъединицы eIF4G и eIF4A, значительно стимулирует терминацию трансляции. Мы показали, что эти субъединицы участвуют в разных стадиях терминации трансляции. eIF4A способствует загрузке eRF1 в A сайт рибосомы, а eIF4G стимулирует активность фактора терминации eRF3. Кроме того, обнаружено, что eIF4A связывается с преТК в АТФ форме и диссоциирует после гидролиза АТФ. Активность eIF4F в терминации трансляции растет в присутствии поли(А) хвоста мРНК и связанного с ним белка PABP. Более того, мы выявили участие в терминации трансляции фактора инициации eIF4B, который связывается с eIF4F в ходе инициации.
Помимо этого, нами обнаружено, что другой фактор инициации, eIF3, также значительно стимулирует терминацию трансляции. eIF3 не только стимулирует распознавание стоп-кодона фактором терминации eRF1, но и активирует следующую стадию терминации – гидролиз ГТФ фактором eRF3 и последующую аккомодацию eRF1 в пептидилтрансферазном центре рибосомы. Показано, что для работы eIF3 в терминации трансляции не требуется eIF4F, и при совместном добавлении этих факторов инициации в реакцию гидролиза пептидил-тРНК их эффекты суммируются. Обнаружено также, что eIF3 связывается с терминационными комплексами, однако не влияет на связывание с ними факторов

терминации. Помимо этого, нами подтверждено участие eIF3 в сквозном чтении преждевременных стоп-кодонов, которое ранее было показано на клетках дрожжей.
Выводы: В ходе исследования было продемонстрировано участие двух больших инициаторных комплексов в регуляции терминации трансляции – комплекса eIF4F, связывающегося с кэпом на мРНК, и комплекса eIF3, связывающегося с 40S субъединицей рибосомы. Белки, входящие в эти комплексы, либо участвуют в образовании closed-loop структуры мРНК, либо могут быть вовлечены в трансляцию коротких uORF, в результате чего они сближаются с терминирующим комплексом. Таким образом, обнаруженное стимулирующее влияние инициаторных факторов на терминацию трансляции происходит, вероятно, при трансляции этих структур.
Финансирование: Исследование поддержано грантом РНФ (№ 22-14-00279).

Regulation of eukaryotic translation termination by initiation complex

Alkalaeva E.Z.*, Shuvalova E.Y., Shuvalov A.V., Bizayev N.S., Al Sheikh W.
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
** alkalaeva@eimb.ru*

Key words: translation termination; translation initiation; ribosome

Motivation and Aim: In addition to the two main release factors eRF1 and eRF3, the eukaryotic termination complex may include many more proteins. Thus, in yeast and mammals, the regulation of translation termination by PABP, eIF3j, DDX19, Gle1, eIF5A, ABCE1, PDCD4, PAIP1 and PAIP2 has been shown. In addition, it is obvious that in a number of processes occurring during translation, such as the formation of a mRNA closed-loop structure and translation at short uORFs with subsequent reinitiation of translation, the initiation and termination ribosomal complexes come closer together. Thus, initiation factors can potentially influence the activity of release factors. The goal of this work was a comprehensive study of the role of initiation factors in eukaryotic translation termination.
Methods and Algorithms: To solve this problem, we used *in vitro* model systems: the mammalian reconstituted translation system from individual components, and pretermination complexes purified from rabbit reticulocyte lysate. Using these methods, the influence of translation initiation factors on different stages of translation termination was studied.
Results: We found that the initiation factor eIF4F, which binds to PABP, or more precisely its two subunits eIF4G and eIF4A, significantly stimulates translation termination. We have shown that these subunits are involved in different stages of translation termination. eIF4A promotes the loading of eRF1 into the A site of the ribosome, and eIF4G stimulates the activity of the release factor eRF3. In addition, eIF4A was found to bind to preTC in the ATP form and dissociate after ATP hydrolysis. The translation termination activity of eIF4F increases in the presence of the poly(A) tail of the mRNA and PABP. Moreover, we identified the involvement of the initiation factor eIF4B in translation termination, which binds to eIF4F during initiation.

We found also that another initiation factor, eIF3, significantly stimulates translation termination. eIF3 not only stimulates the recognition of the stop codon by eRF1, but also activates the next stage of termination – GTP hydrolysis by eRF3 and subsequent accommodation of eRF1 in the peptidyl transferase center of the ribosome. It has been shown that eIF3 does not require eIF4F to function in translation termination, and when these initiation factors are added together to the peptidyl-tRNA hydrolysis, their effects are additive. It was also found that eIF3 binds to termination complexes, but it does not affect the binding of release factors to them. In addition, we confirmed the participation of eIF3 in the readthrough of premature stop codons, which was previously shown in yeast cells.

Conclusion: The study demonstrated the involvement of two large initiation complexes in the regulation of translation termination: the eIF4F complex, which binds to capped mRNA, and the eIF3 complex, which binds to the 40S ribosomal subunit. Proteins included in these complexes either participate in the formation of the closed-loop structure of mRNA or can be involved in the translation of short uORFs, as a result of which they come closer to the termination complex. Thus, the detected stimulating effect of initiation factors on translation termination probably occurs during translation of these structures.

Funding: The study is supported by RSF (No. 22-14-00279).

doi 10.18699/bgrs2024-11.2-02

Последний смысловой кодон и поли(А)-хвост мРНК регулируют терминацию трансляции эукариот

Бизяев Н.*, Шувалов А., Салман А., Егорова Т., Шувалова Е., Алкалаева Е.

Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва, Россия

** bizyaev@eimb.ru*

Ключевые слова: трансляция; терминация; стоп-кодон; сквозное прочтение; факторы терминации; регуляция трансляции; поли(А)-хвост; контекст мРНК

Мотивация и цель: Когда рибосома достигает стоп-кодона, то обычно происходит завершение синтеза белка – терминация трансляции. Однако за счет распознавания стоп-кодона частично комплементарной тРНК на стоп-кодоне может происходить продолжение элонгации трансляции, т. е. сквозное прочтение стоп-кодона. Вероятно, конкуренция терминации трансляции и сквозного прочтения обусловлена влиянием окружения, в котором находится стоп-кодон. Особенно сильное влияние такого окружения стоп-кодона на терминацию трансляции обнаружено у некоторых инфузорий, в генетическом коде которых все три стоп-кодона могут кодировать аминокислоты. Для предсказания событий, происходящих на стоп-кодоне, Л.Л. Киселевым была предложена гипотеза о существовании «третьего генетического кода» [1]. В рамках развития этой концепции мы изучили, как такие структурные элементы мРНК, как последний смысловой кодон и поли(А)-хвост, влияют на эффективность эукариотической терминации трансляции. Ранее для бактерий было показано, что на стоп-кодонах, следующих за кодонами пролина и глицина, наблюдается значительное понижение эффективности терминации трансляции [2]. Однако для эукариот этот вопрос остается неизученным. Роль поли(А)-хвоста в инициации трансляции и стабильности мРНК хорошо исследована. Кроме того, в нашей лаборатории было показано, что белок, связывающийся с поли(А)-хвостом (PABP), стимулирует терминацию трансляции [3]. Полученные нами данные о влиянии этих элементов мРНК на терминацию трансляции указывают на то, что обе эти структуры являются важными регуляторами терминации трансляции. Дальнейшее систематическое изучение механизмов, лежащих в основе их работы, позволит по последовательности мРНК предсказывать эффективность протекания терминации трансляции на стоп-кодонах.

Методы и алгоритмы: Для изучения влияния 5′-контекста стоп-кодона на эффективность терминации трансляции мы использовали очищенные рибосомные комплексы, остановленные на стоп-кодоне (претерминационные комплексы), которые были собраны на мРНК с различными кодонами, предшествующими стоп-кодонам, и длиной поли(А)-хвоста. Такие комплексы были использованы для оценки эффективности высвобождения синтезируемого белка или пептида в ходе терминации трансляции. Высвобождение в ходе терминации трансляции длинного белка изучали с помощью Termi-Luc анализа [4], тогда как высвобождение короткого пептида, моделирующего этот процесс на короткой рамке считывания (uORF), – в реконструированной системе эукариотической трансляции [5].

Результаты: Мы обнаружили, что последний смысловой кодон в кодирующей последовательности может определять скорость эукариотической терминации трансляции белка. Так, при высвобождении длинного пептида скорость

терминации была снижена на глициновых кодонах. А на пролиновых кодонах у эукариот, в отличие от бактерий, скорость терминации трансляции была снижена только на одном из двух изученных кодонов. Интересно, что, кроме сниженной скорости терминации трансляции, на всех кодонах глицина было обнаружено уменьшение стабильности претерминационного комплекса. Таким образом, кодон глицина перед стоп-кодоном не является оптимальным для реакции терминации трансляции длинного белка. Однако в другой системе, где тестировали высвобождение короткого пептида, мы не наблюдали влияния глицина и пролина как последних смысловых кодонов на скорость терминации трансляции. Ранее, при изучении влияния поли(А)-хвоста на эффективность терминации трансляции было обнаружено, что вместе с его удлинением увеличивается эффективность терминации. В рамках данной работы нам удалось установить, что на короткой кодирующей последовательности присутствие поли(А)-хвоста не стимулировало терминацию трансляции, если последним смысловым кодоном были кодоны глицина или пролина.

Выводы: Последовательность последнего смыслового кодона и длина поли(А)-хвоста являются контекстными детерминантами эффективности эукариотической терминации трансляции. Эффекты этих детерминант также модулируются дополнительными факторами, такими как присутствие PABP и длина синтезируемого пептида. Это указывает на важность комплексного понимания локального окружения, в котором находится стоп-кодон, для успешного предсказания, продолжится или завершится на нем синтез белка.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 23-24-00382, https://rscf.ru/project/23-24-00382/.

The last sense codon and the poly(A) tail of mRNA regulate the termination of eukaryotic translation

Biziaev N.*, Shuvalov A., Salman A., Egorova T., Shuvalova E., Alkalaeva E.

Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia

* *bizyaev@eimb.ru*

Key words: translation; termination; stop codon; readthrough; release factors; translational control; poly(A) tail; mRNA context

Motivation and Aim: When a ribosome reaches a stop codon, protein synthesis usually ends, which is called translation termination. However, in certain cases, translation elongation may continue at the stop codon due to the recognition of the stop codon by a partially complementary tRNA, a phenomenon referred to as readthrough. The competition between translation termination and readthrough is likely to be influenced by the surrounding context of the stop codon. Notably, in some ciliates, where all three stop codons can encode amino acids, the context surrounding stop codons exerts a particularly strong influence. To predict events occurring at the stop codon, L.L. Kiselev proposed a hypothesis about the existence of a “third genetic code” [1]. As part of the development of this concept, we studied how such structural elements of mRNA as the last sense codon and the poly(A) tail affect the efficiency of eukaryotic translation termination. While previous research in bacteria has demonstrated a significant decrease in translation termination efficiency at stop codons following proline and glycine codons [2], this phenomenon remains unexplored in eukaryotes. Furthermore, the role of the poly(A) tail in translation initiation and mRNA stability has been extensively studied. In

addition, our laboratory has shown that poly(A)-binding protein (PABP) stimulates translation termination [3]. Our data indicate that both the last sense codon and the poly(A) tail serve as important regulators of translation termination in eukaryotes. Further systematic investigation into the mechanisms underlying the functioning of mRNA structural elements in translation termination will enable the prediction of translation termination efficiency at stop codons based on mRNA sequence.

Methods and Algorithms: To study the effect of the 5′ stop codon context on translation termination efficiency, we used purified ribosomal complexes frozen at the stop codon (pretermination complexes). These complexes were assembled on mRNAs with varying last sense codons and poly(A) tail lengths. Such complexes were used to evaluate the release efficiency of the synthesized protein or peptide during translation termination. For the analysis of the release of a long protein, we utilized Termi-Luc analysis [4] while the release of a short peptide, mimicking this process on a short reading frame (uORF), was studied within a reconstructed eukaryotic translation system [5].

Results: We found that the last sense codon within a coding sequence can determine the rate of eukaryotic termination of protein translation. Specifically, we observed a reduction in termination rate at glycine codons upon the release of a long peptide Furthermore, unlike in bacteria, where translation termination efficiency is reduced at any proline codon, in eukaryotes, this reduction was observed only at one of the two studied proline codons. Interestingly, in addition to the diminished termination rate, there was a decrease in the stability of the pretermination complex at all glycine codons. Thus, the glycine codon before the stop codon is not optimal for the translation termination reaction of a long protein. However, in another experimental system focusing on the release of a short peptide, we did not observe any effect of glycine and proline as the last sense codons on the rate of translation termination. Furthermore, our previous studies on the effect of the poly(A) tail on translation termination efficiency indicated that the length of the poly(A) tail correlates with increased termination efficiency. Our current work demonstrates that, in the context of a short coding sequence, the presence of a poly(A) tail does not stimulate translation termination when the last sense codon is either glycine or proline.

Conclusion: The sequence of the last sense codon and the length of the poly(A) tail emerge as contextual determinants influencing the efficiency of eukaryotic translation termination. Moreover, these determinants are subject to modulation by additional factors, including the presence of PABP and the length of the synthesized peptide. This indicates the importance of a comprehensive understanding of the local context surrounding a stop codon to accurately predict whether protein synthesis will persist or terminate at that point.

Funding: This work was supported by the Russian Science Foundation (Grant No. 23-24-00382), https://rscf.ru/project/23-24-00382/.

Список литературы/References

1. Kiselev L.L. The third genetic code. *Priroda*. 2006;9:3-9 (in Russian)
2. Pierson W.E. et al. Uniformity of Peptide Release Is Maintained by Methylation of Release Factors. *Cell Rep*. 2016;17(1):11-18
3. Ivanov A. et al. PABP enhances release factor recruitment and stop codon recognition during translation termination. *Nucleic Acids Res*. 2016;44(16):7766-7776
4. Susorov D., Egri S., Korostelev A.A. Termi-Luc: a versatile assay to monitor full-protein release from ribosomes. *RNA*. 2020;26(12):2044-2050
5. Alkalaeva E.Z. et al. *In vitro* reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell*. 2006;125(6):1125-1136

doi 10.18699/bgrs2024-11.2-03

Особенности взаимодействия piRNA с mRNA гена *FGF2*

Иващенко А.[1, 2]*, Пыркова А.[1, 2], Акимниязова А.[1], Керимов Т.[1], Ершов И.[1], Султанбеков С.[1]

[1] *КазНУ им. Аль-Фараби, Алматы, Казахстан*

[2] *ТОО «Центр биоинформатики и наномедицины», Алматы, Казахстан*

* *a.iavashchenko@gmail.com*

Ключевые слова: ген; *FGF2*; mRNA; piRNA; человек

Мотивация и цель: Семейство генов факторов роста фибробластов (*FGF*) состоит из 23 генов. Многие из них, в том числе *FGF2,* участвуют в развитии различных заболеваний. Поэтому представляется важным выяснить, могут ли piRNA участвовать в регуляции экспрессии этих генов [1]. Нами установлено, что mRNA некоторых генов семейства *FGF* не связываются с piRNA, в то время как mRNA гена *FGF2* взаимодействовала со многими piRNA и важно выявить особенности их связывания.

Методы и алгоритмы: Программа MirTarget определяет начало сайтов связывания piRNA и mRNA; локализацию сайтов связывания в 5'UTR, CDS и 3'UTR; свободная энергия гибридизации (ΔG, кДж/моль), величина ΔG/ΔGm (%), где ΔGm равна свободной энергии связывания piRNA с полностью комплементарной нуклеотидной последовательностью и схемой взаимодействия нуклеотидов piRNA-mRNA. Выборочно были отобраны сайты связывания piRNA-mRNA со значением ΔG/ΔGm, равным 90 % и более. Положение сайтов связывания указано от первого нуклеотида в 5'UTR mRNA. Взаимодействие нуклеотидов между piRNA и mRNA генов-мишеней не только в аденине и урациле (A-U), гуанине и цитозине (G-C), но и между A-C и G-U.

Результаты: Установлено, что из одного миллиона piRNA (piR-10 to piR-999999) с mRNA гена *FGF2* связывались piRNA только в 3'UTR имеющей длину 6652 nt. С mRNA гена *FGF2* связывались 219 piRNA начиная с piR-34640 до piR-950352. Из них piR-70255, piR-197381, piR-197597, piR-200742, piR-207197, piR-293655, piR-325736, piR-417654, piR-436961, piR-520416 и piR-919763 имели длину 34 нт. Для piR-618251 длиной 31 нуклеотид приведена схема взаимодействия с 3'UTR mRNA гена *FGF2* (рис.1) начиная с позиции 2910 нт. piR-618251 взаимодействовала с mRNA за счет образования 28 канонических пар A-U, G-C и трех неканонических пар A-C и G-U. Величина свободной энергии связывания равная –161 kJ/mol составляла 94 % от ΔG/ΔGm – максимальной свободной энергии связывания piR-618251 с mRNA.

Все сайты связывания piRNA располагались в 3'UTR участке mRNA с 2836 до 3134 нт, то есть длиной 298 нт. Распределение сайтов связывания piRNA в этом участке было с отчетливым максимумом в центре участка. Были выявлены piRNA начало сайтов связывания которых совпадали, и они образовывали группы piRNA от двух piRNA до 21 piRNA имеющих начало сайтов связывания в одной позиции. В сайтах связывания с началом с 2888 и с 2889 нт связывались по пять piRNA. В сайтах с началом связывания с 2892, 2893, 2897 и 2898 нт связывались по шесть piRNA, На рис. 2 показано как 21 piRNA могут связываться с mRNA гена *FGF2* в

позиции 2910 нт. Число неканонических пар связывания (G-U и A-C) изменялось от одной до четырех пар, что соответствует величине ΔG/ΔGm более 90 %.

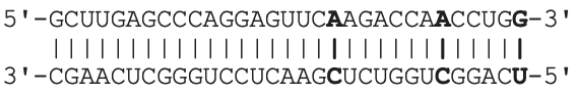


Рис. 1. Схема взаимодействия piR-618251 с mRNA гена *FGF2*. Жирным шрифтом выделены неканонические пары нуклеотидов

```
5'-GCUUGAGCCCAGGAGUUCAAGACCAACCUGG-3' mRNA FGF2
3'-CGAACUCGGGUCCUCAAGCUCUGGUCGGACU-5' piR-618251
3'-CGAGCUCGAGUCUUCAAGUUCUGGUCGGAC-5'  piR-723102
3'-CGAGUUCAAGUCCUCAAGUUCUGGUCGGA-5'   piR-591009
3'-CGAACUUGAGUCCUCAAGCUCUGGUCGGA-5'   piR-635311
3'-CGAACUCAGGUCCUCAAGUUCUGGUCAG-5'    piR-414978
3'-CGAAUUCAAGUCCUCAAGUUCUGGUCG-5'     piR-811352
3'-CGAACUCAAGUUCUCAAGUUCUGGUC-5'      piR-42935
3'-CGAAUUCGAGUCCUCAAGUUCUGGUU-5'      piR-51132
3'-CGAACUCGGGUCCUCAAGUUUUGGUU-5'      piR-114311
3'-CGAACUCGGGUCCUCAAGUCCUGGUC-5'      piR-613002
3'-CGAACUCAGGUCCUUAAGCUCUGGUC-5'      piR-711504
3'-UGGACUCGGACCCUCAAGUUCUGGU-5'       piR-283642
3'-CGAACUCGAGUCCUCAAGUUCUAGU-5'       piR-411668
3'-CGAACUCAAGUUCUCAAGUUCUGGU-5'       piR-534739
3'-CGAACUCGAGUCCUCAAGUUCUGGU-5'       piR-555444
3'-CGGACUCGAGUCCCCAAGUUCUGGU-5'       piR-593416
3'-CGAACUCGAGUCCUCAAGUUCUAAU-5'       piR-708359
3'-CGGACUCGGGUCCUCAAGCUUUGGU-5'       piR-814823
3'-CGAACUCAGGUCCUCAAGUUCCGA-5'        piR-87465
3'-CGAAUUUGGGUCCUCAAGUUCUGG-5'        piR-230717
3'-CGAACUCAAGUCCUUAAGUUCUGG-5'        piR-441972
```

Рис. 2. Нуклеотидные последовательности piRNA связывающихся с позиции 2910 нт кластера сайтов связывания в mRNA гена *FGF2*. Жирным шрифтом выделены нуклеотиды неканонических пар

Весь участок mRNA длиной 273 нт фактически является одним кластером сайтов связывания 219 piRNA. Полученные результаты свидетельствует о конкуренции этих piRNA за связывание с mRNA гена *FGF2*. С учетом свободной энергии взаимодействия и концентрации piRNA можно предсказывать какой эффект будет наблюдаться для каждой piRNA. Кроме piRNA из данной группы нужно учитывать и другие piRNA которые тоже конкурентно могут связываться с mRNA если их сайты связывания перекрываются с сайтами связывания piRNA этой группы и других групп piRNA. В настоящее время трудно интерпретировать зачем столько piRNA вовлечены в контроль экспрессии одного гена. При этом сайты связывания

этих piRNA расположены компактно в небольшом участке mRNA. Учитывая, что данный пример не является единственным, и наблюдается для других ассоциаций miRNA и piRNA с mRNA [2–6], то обнаруженные явления требуют дальнейшего исследования.
Выводы: piRNA могут подавлять трансляцию mRNA гена *FGF2* путем образования канонических и неканонических пар нуклеотидов.

Features of the interaction of piRNA with mRNA of the *FGF2* gene

Ivashchenko A.[1, 2]*, Pyrkova A.[1,2], Akimniyazova A.[1], Kerimov T.[1], Yershov I.[1], Sultanbekov S.[1]
[1] *al-Farabi Kazakh National University, Almaty, Kazakhstan*
[2] *Center for Bioinformatics and Nanomedicine, Almaty, Kazakhstan*
* *a.iavashchenko@gmail.com*

Key words: gene; *FGF2*; mRNA; piRNA; human

Motivation and aim: The fibroblast growth factor (*FGF*) gene family consists of 23 genes. Many of them, including FGF2, are involved in the development of various diseases. Therefore, it seems important to find out whether piRNAs can participate in the regulation of the expression of these genes [1]. We found that the mRNA of some genes of the *FGF* family does not bind to piRNA, while the mRNA of the *FGF2* gene interacted with many piRNA and it is important to identify the features of their binding.
Methods and algorithms: MirTarget program defines the beginning of piRNA and mRNA binding sites; localization of binding sites in the 5'UTR, CDS and 3'UTR; free energy of hybridization (ΔG, kJ/mol), the ΔG/ΔGm (%) value, where ΔGm equals the free binding energy of piRNAs with fully complementary nucleotide sequence and scheme of piRNA-mRNA nucleotides interaction. There were selectively taken the piRNAs-mRNAs binding sites with ΔG/ΔGm value is equal to 90 % and more. Position of binding sites indicated from the first nucleotide in 5'UTR of mRNA. The interaction of nucleotides between piRNAs and mRNAs of target genes not only in adenine and uracil (A-U), guanine and cytosine (G-C), but between A-C and G-U.
Results: It was found that out of one million piRNA (piR-10 to piR-999999) the mRNA of *FGF2* gene was bound by piRNA only in the 3'UTR, which is 6652 nt in length. 219 piRNAs, starting from piR-34640 to piR-950352, were associated with FGF2 gene mRNA. Of these, piR-70255, piR-197381, piR-197597, piR-200742, piR-207197, piR-293655, piR-325736, piR-417654, piR-436961, piR-520416 and piR-919763 had a length 34 nt. For piR-618251, 31 nucleotides long, a diagram of interaction with the 3'UTR mRNA of the *FGF2* gene is shown (Fig. 1) starting from position 2910 nt. piR-618251 interacted with mRNA through the formation of 28 canonical A-U, G-C pairs and three non-canonical A-C and G-U pairs. The free energy of binding equal to –161 kJ/mol was 94 % of ΔG/ΔGm – the maximum free energy of binding of piR-618251 to mRNA.
All piRNA binding sites were located in the 3'UTR region of mRNA from 2836 nt to 3134 nt, i.e., 298 nt in length. The distribution of piRNA binding sites in this region was with a distinct maximum in the center of the region. piRNAs whose binding site origins

coincided were identified, and they formed piRNA groups from two piRNAs to 21 piRNAs with binding site origins in the same position. At the binding sites starting at 2888 nt and at 2889 nt, five piRNAs were bound. At the binding sites at 2892 nt, at 2893 nt, at 2897 nt, and at 2898 nt, six piRNAs were bound. Figure 2 shows how 21 piRNAs can bind to the FGF2 gene mRNA at position 2910 nt. The number of non-canonical binding pairs (G-U and A-C) varied from one to four pairs, which corresponds to a ΔG/ΔGm value of more than 90 %.

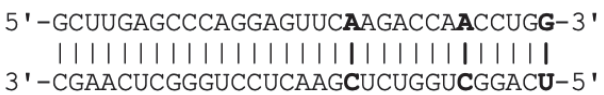


Fig. 1. Scheme of interaction of piR-618251 with mRNA of *FGF2* gene. Non-canonical nucleotide pairs are highlighted in bold

```
5'-GCUUGAGCCCAGGAGUUCAAGACCAACCUGG-3' mRNA FGF2
3'-CGAACUCGGGUCCUCAAGCUCUGGUCGGACU-5' piR-618251
3'-CGAGCUCGAGUCUUCAAGUUCUGGUCGGAC-5'  piR-723102
3'-CGAGUUCAAGUCCUCAAGUUCUGGUCGGA-5'   piR-591009
3'-CGAACUUGAGUCCUCAAGCUCUGGUCGGA-5'   piR-635311
3'-CGAACUCAGGUCCUCAAGUUCUGGUCAG-5'    piR-414978
3'-CGAAUUCAAGUCCUCAAGUUCUGGUCG-5'     piR-811352
3'-CGAACUCAAGUUCUCAAGUUCUGGUC-5'      piR-42935
3'-CGAAUUCGAGUCCUCAAGUUCUGGUU-5'      piR-51132
3'-CGAACUCGGGUCCUCAAGUUUUGGUU-5'      piR-114311
3'-CGAACUCGGGUCCUCAAGUCCUGGUC-5'      piR-613002
3'-CGAACUCAGGUCCUUAAGCUCUGGUC-5'      piR-711504
3'-UGGACUCGGACCCUCAAGUUCUGGU-5'       piR-283642
3'-CGAACUCGAGUCCUCAAGUUCUAGU-5'       piR-411668
3'-CGAACUCAAGUUCUCAAGUUCUGGU-5'       piR-534739
3'-CGAACUCGAGUCCUCAAGUUCUGGU-5'       piR-555444
3'-CGGACUCGAGUCCCCAAGUUCUGGU-5'       piR-593416
3'-CGAACUCGAGUCCUCAAGUUCUAAU-5'       piR-708359
3'-CGGACUCGGGUCCUCAAGCUUUGGU-5'       piR-814823
3'-CGAACUCAGGUCCUCAAGUUCCGA-5'        piR-87465
3'-CGAAUUUGGGUCCUCAAGUUCUGG-5'        piR-230717
3'-CGAACUCAAGUCCUUAAGUUCUGG-5'        piR-441972
```

Fig. 2. Nucleotide sequences of piRNA binding from position 2910 nt of the binding site cluster in the mRNA of *FGF2* gene. Nucleotides of non-canonical pairs are highlighted in bold

The entire 273-nt region of mRNA is actually one cluster of 219 piRNA binding sites. The results obtained indicate that these piRNAs compete for binding to the *FGF2* gene mRNA. Considering the free energy of interaction and the concentration of piRNA, it is possible to predict what effect will be observed for each piRNA. In addition to piRNA from this group, it is necessary to consider other piRNAs that can also competitively

bind to mRNA if their binding sites overlap with the binding sites of piRNAs of this group and other piRNA groups. At present, it is difficult to interpret why so many piRNAs are involved in controlling the expression of a single gene. Moreover, the binding sites for these piRNAs are located compactly in a small region of mRNA. Considering that this example is not the only one and is observed for other associations of miRNA and piRNA with mRNA [2-6], the detected phenomena require further study. *Conclusion*: piRNAs can suppress the translation of *FGF2* gene mRNA by forming canonical and non-canonical nucleotide pairs.

Список литературы/References

1. Wang J., Shi Y., Zhou H. et al. piRBase: integrating piRNA annotation in all aspects. *Nucleic Acids Res.* 2021;50:265-272. doi 10.1093/nar/gkab1012
2. Rakhmetullina A., Akimniyazova A., Niyazova T., Pyrkova A., Ivashchenko A., Zielenkiewicz P. Endogenous piRNAs Can Interact with the Omicron Variant of the SARS-CoV-2 Genome. *Curr Issues Mol Biol.* 2023;45(4):2950-2964
3. Kamenova S., Sharapkhanova A., Akimniyazova A., Pyrkova A., Ivashchenko A. piRNA and miRNA Can Suppress the Expression of Multiple Sclerosis Candidate Genes. *Nanomaterials.* 2023;13(1):22
4. Akimniyazova A., Niyazova T., Yurikova O., Zhanuzakov M., Ivashchenko A. piRNAs may regulate expression of candidate genes of esophageal adenocarcinoma. *Front Genet*. 2022;13:1069637
5. Akimniyazova A., Yurikova O., Pyrkova A., Ryskulova A.-G., Ivashchenko A. In Silico Study of piRNA Interactions with the SARS-CoV-2 Genome. *Int J Mol Sci.* 2022;23(17):9919
6. Rakhmetullina_A., Akimniyazova A., Niyazova T., Pyrkova A., Tauassarova_M., Ivashchenko A., Zielenkiewicz P. Interactions of piRNAs with the mRNA of Candidate Genes in Esophageal Squamous Cell Carcinoma. *Curr Issues Mol Biol*. 2023;45(7):6140-6153. doi 10.3390/cimb45070387

doi 10.18699/bgrs2024-11.2-04

РНК-связывающие свойства многофункционального белка NUCB1 и его возможная роль в регуляции трансляции

Костарева О.*, Михайлина А., Тищенко С.
Институт белка РАН, Пущино, Россия
** kostareva@vega.protres.ru*

Ключевые слова: нуклеобиндин 1; NUCB1; микро-РНК; регуляция трансляции

Мотивация и цель: Нуклеобиндин 1 (NUCB1) – консервативный Ca^{2+}-связывающий эукариотический белок. NUCB1 опосредованно вовлечен в регуляцию множества процессов в организме в результате его взаимодействия различными партнёрами: белками: Gαi [1], Hsc70-interacting protein [2], APP [3]**;** липидами [4]; ДНК [5, 6]. Предполагается, что ДНК-связывающая активность NUCB1 может играть роль индуктора аутоиммунных реакций [7]. Было показано, что NUCB1 в качестве транскрипционного фактора [6] принимает участие в активации эпителиально-мезенхимального перехода, связываясь с последовательностью ДНК E-box *cripto* промотора. Недавно мы впервые показали, что NUCB1 способен образовывать прочные комплексы с некоторыми микроРНК человека, ассоциированными с развитием системной красной волчанки и эпителиально-мезенхимальным переходом [8]. Мы предполагаем, что взаимодействие белка NUCB1 с РНК имеет функциональное значение и может влиять на регуляцию экспрессии генов на уровне трансляции. Целью настоящей работы было исследование роли доменов NUCB1 во взаимодействии с РНК.
Методы и алгоритмы: Было получено 5 фрагментов белка NUCB1, содержащих разные домены белка, РНК-связывающая способность трёх фрагментов была исследована с помощью метода поверхностного резонанса плазмонов.
Результаты: NUCB1 образует функциональный димер и состоит из нескольких доменов: сигнальный пептид, ДНК-связывающий домен, Ca^{2+}-связывающий домен (содержит 2 EF-Hand мотива) и лейциновая «застёжка-молния». NUCB1 был впервые обнаружен как фактор роста и дифференцировки В-клеток при системной красной волчанке (СКВ) в культуре клеток селезенки мыши KLM1-7. Было показано, что NUCB1 способен взаимодействовать с тотальной ДНК, выделенной из этой культуры клеток. Систематические инъекции NUCB1 приводили к развитию у модельных линий мышей (MRL/lpr и MRL/n) СКВ-подобного заболевания (появление аутоантител к одно- и двуцепочечной ДНК). Мутантная форма NUCB1, лишенная предположительных участков взаимодействия с ДНК (ДНК-связывающий домен или лейциновая «застёжка-молния»), не стимулировала появления антител к ДНК. Эти данные позволили предположить, что ДНК-связывающая активность NUCB1 играет роль индуктора аутоиммунных реакций. Таким образом, для взаимодействия с ДНК необходимы как ДНК-связывающий домен белка, так и домен лейциновая «застежка-молния». Однако, какова роль доменов во взаимодействии с РНК, до сих пор не было показано. Мы получили три мутантные формы NUCB1, которые содержали кальций-связывающий домен: NUCB1ΔN (239–461 а. о., NUCB1, лишенный ДНК-связывающего домена), NUCB1ΔZ (31–320 а. о., NUCB1, лишенный лейциновой

«застёжки-молнии») и CaBD (228–326 а. о., содержащий только кальций-связывающий домен). Полноразмерный белок NUCB1 и мутантные формы белка были продуцированы в клетках *E.coli* и очищены с помощью хроматографических методов. Мутантная форма белка NUCB1ΔN так же, как полноразмерный белок, формировала гомодимер, а мутантные формы NUCB1ΔZ и CaBD присутствовали в растворе в виде мономеров из-за отсутствия домена лейциновая «застёжка-молния». Мы определили аффинность взаимодействия всех полученных форм белка с РНК методом поверхностного резонанса плазмонов (табл. 1). Показано, что белок NUCB1 и его мутантные формы обладают высоким (наномолярным) и специфическим сродством к РНК, в том числе к микроРНК, ассоциированной с развитие СКВ у человека. Таким образом, было показано, что для взаимодействия с РНК достаточен кальций-связывающий домен белка NUCB1. Однако, несмотря на вовлеченность кальций-связывающего домена, взаимодействие белка с РНК не зависело от присутствия ионов кальция в растворе.

Таблица 1. Константы диссоциации комплексов белка NUCB1 и его мутантных форм с РНК

РНК	K_D, nM			
	NUCB1	NUCB1ΔN	NUCB1ΔZ	CaBD
miR-200a-3p	3.8 ± 0.1	0.6 ± 0.09	0.3 ± 0.01	5.2 ± 0.2
Олиго (C)	Взаимодействие не детектировалось			
Олиго (A)	1.2 ± 0.1	1.4 ± 0.1	0.3 ± 0.02	5.7 ± 0.5
Олиго (U)	1.1 ± 0.1	1.7 ± 0.3	0.4 ± 0.03	0.2 ± 0.04
Олиго (G)	1.5 ± 0.2	5.3 ± 0.6	3.1 ± 0.2	18.2 ± 3

Для того чтобы оценить необходимость участия CaBD во взаимодействии белка с РНК, мы создали две мутантные формы белка, не содержащие CaBD – NUCB1 DBD (31–220 а. о.) и NUCB1 LZ (328–416 а. о.). В настоящее время ведётся исследование РНК-связывающей способности этих мутантных форм.

Выводы: Показано, что кальций-связывающий домен белка NUCB1 достаточен для связывания белка с РНК, причём, это взаимодействие является кальций-независимым. В литературе описан только один случай вовлеченности EF-hand мотива белка VILIP в связывание с РНК [9], но это взаимодействие зависит от ионов кальция. По-видимому, белок NUCB1 узнает РНК по уникальному, пока не исследованному, механизму и может опосредованно участвовать в регуляции трансляции РНК посредством взаимодействия с микроРНК.

RNA-Binding capacity of multifunctional NUCB1 protein and its proposal role in translation regulation

Kostareva O.*, Mikhaylina A., Tishchenko S.

Institute of Protein Research, Russian Academy of Sciences, Pushchino, Rossiya

** kostareva@vega.protres.ru*

Key words: nucleobindin 1; NUCB1; micro-RNA; translation regulation

Motivation and Aim: Nucleobindin 1 (Calnuc, NUCB1) is a highly conserved multi-domain DNA- and calcium-binding eukaryotic protein. NUCB1 is involved in the

regulation of many processes via interactions with proteins: Gαi [1], carboxyl terminus of Hsc70-interacting protein [2], APP [3]; lipids [4], and DNA [5,6]. DNA-binding ability is suggested to can play the role autoimmune reactions inductor [7]. It has been shown that NUCB1 is involved in epithelial-mesenchymal transition (EMT) activation as a transcription factor through binding to the canonical enhancer box (E-box) sequence of the Cripto-1 gene promoter. Recently we have shown that NUCB1 is capable to form high-affinity complexes with some of human micro-RNAs which are associated with system lupus erythematosus (SLE) and EMT [8]. We assume that RNA-binding ability of NUCB1 has the functional significance and could be develop on gene expression regulation at the translational level. The purpose of current study was investigation of contribution of different NUCB1 domains to the RNA interaction.

Methods and Algorithms: Five truncated forms of NUCB1 protein have been produced and RNA-binding ability for three of them has been estimated by surface plasmon resonance technique.

Results: NUCB1 forms functional dimer and consists of several independent functional domains: a signal peptide, a basic amino acid-rich DNA-binding domain (DBD), calcium-binding domain (contains two Ca2+-binding EF-hand motifs) (CaBD), a C-terminal leucine zipper (LZ). NUCB1 was first found as B-cells growth and differentiation factor in spleen cell culture KLM1-7 of SLE-prone mice. NUCB1 has been shown to bind total DNA extracted from these cells. Systematic injections of NUCB1 into SLE-prone mice (MRL/lpr и MRL/n) resulted in increased production of anti-ssDNA and anti-dsDNA autoantibodies. Injections of mutant forms of NUCB1 lacked proposal fragments for DNA interaction (DBD and LZ) didn't result in increased production anti-DNA autoantibodies. Thus, one may assume a potential role of NUCB1 DNA-binding capacity as inductor of autoimmune processes. It was established that the DBD and the LZ domain are important for DNA binding, but their involvement in NUCB1 binding to RNA has never been shown. We obtained three mutant forms of NUCB1 all of which contains CaBD: NUCB1ΔN (239–461 aa, NUCB1, lacked DBD), NUCB1ΔZ (31–320 aa, NUCB1, lacked LZ) and CaBD (228–326 aa, containing EF-hand motifs). NUCB1 and its mutant forms were produced in *E. coli* cells and purified using chromatographic techniques. The truncated forms NUCB1ΔZ and CaBD were monomers due to the absence of LZ, whereas NUCB1ΔN formed dimers similarly to NUCB1. We have defined affinity of all produced mutant forms to RNA by surface plasmon resonance technique (Table 1). Apparently, all truncated NUCB1 variants retained high (nanomolar) affinity for RNA. They are also bind RNA in specific manner, including to micro-RNA associated with SLE in human. CaBD alone has been shown to be sufficient for RNA binding. Calcium ions did not affect the binding of the truncated NUCB1 forms to RNAs.

Table 1. The K_D (nM) of complexes formed by the NUCB1 mutant forms with RNAs

РНК	K_D, nM			
	NUCB1	NUCB1ΔN	NUCB1ΔZ	CaBD
miR-200a-3p	3.8 ± 0.1	0.6 ± 0.09	0.3 ± 0.01	5.2 ± 0.2
Oligo (C)	Not detected			
Oligo (A)	1.2 ± 0.1	1.4 ± 0.1	0.3 ± 0.02	5.7 ± 0.5
Oligo (U)	1.1 ± 0.1	1.7 ± 0.3	0.4 ± 0.03	0.2 ± 0.04
Олиго (G)	1.5 ± 0.2	5.3 ± 0.6	3.1 ± 0.2	18.2 ± 3

To estimate necessary of CaBD into RNA-interaction of NUCB1 we produced two additional NUCB1 truncated forms lacked CaBD – NUCB1 DBD (31–220 aa) and NUCB1 LZ (328–416 aa). We are studying their RNA-binding ability now.

Conclusion: Calcium-binding domain of NUCB1 protein is shown to sufficient to the interaction with RNA, moreover the interaction is calcium-independently. Only one article is devoted to involvement of EF-hand motif of VILIP protein in RNA-binding [9], but that interaction is calcium-dependent. Apparently, NUCB1 protein recognize RNA using a unique, not yet studied mechanism, and could be involved in the regulation of translation of mRNAs via interaction with microRNAs.

Список литературы/References

1. Kapoor N., Gupta R., Menon S.T. et al. Nucleobindin 1 is a calcium-regulated guanine nucleotide dissociation inhibitor of G_i1. *J Biol Chem*. 2010;285:31647-31660. doi 10.1074/jbc.M110.148429
2. Xue F., Wu Y., Zhao X. et al. CHIP mediates down-regulation of nucleobindin-1 in preosteoblast cell line models. *Cell Signal*. 2016;28:1058-1065. doi 10.1016/j.cellsig.2016.04.016
3. Lin P., Li F., Zhang Y.-W. et al. Calnuc binds to Alzheimer's beta-amyloid precursos protein and affects its biogenesis. *J Neurochem*. 2007;100:1505-1514. doi 10.1111/j.1471-4159.2006.04336.x
4. Niphakis M.J., Lum K.M., Cognetta A.B. et al. A global map of lipid-binding proteins and their ligandability in cells. *Cell*. 2015;16:1668-1680. doi 10.1016/j.cell.2015.05.045
5. Miura K., Titani K., Kurosawa Y., Kanai Y. Molecular cloning of nucleobindin, a novel DNA-binding protein that contains both a signal peptide and a leucine zipper structure. *Biochem Biophys Res Commun*. 1992;187:375-380. doi 10.1016/s0006-291x(05)81503-7
6. Sinha S., Pattnaik S., Aradhyam G.K. Molecular evolution guided functional analyses reveals Nuckeobindin-1 as a canonical E-box binding protein promoting Epithelial-to-Mesenchymal transition (EMT). *Biochim Biophys Acta Proteins Proteom*. 2019;1867:765-775. doi 10.1016/j.bbapap.2019.05.009
7. Kanai Y., Takeda O., Kanai Y., Miura K., Kurosawa Y. Novel autoimmune phenomena induced *in vivo* by a new DNA binding protein Nuc: A study on MRL/n mice. *Immunol Lett*. 1993;39:83-89. doi 10.1016/0165-2478(93)90168-2
8. Mikhaylina A., Svoeglazova A., Stolboushkina E., Tishchenko S., Kostareva O. The RNA-binding and RNA-melting activities of the multifunctional protein nucleobindin 1. *Int J Mol Sci*. 2023;24:6193. doi 10.3390/ijms24076193
9. Mathisen P.M., Johnson J.M., Kawczak J.A., Tuohy V.K. Visinin-like protein (VILIP) is a neuron-specific calcium-dependent double-stranded RNA-binding protein. *J Biol Chem*. 1999;274:31571-31576. doi 10.1074/jbc.274.44.31571

doi 10.18699/bgrs2024-11.2-05

Regulatory functions of the upstream open reading frame in mRNA of the *MCRS1* human gene

Baltin S.M.[1]*, Suspitsyna A.D.[2], Shepelev N.M.[3, 4], Rubtsova M.P.[3, 4], Dontsova O.A.[2, 3, 4]

[1] *Biological Faculty, Lomonosov Moscow State University, Moscow, Russia*

[2] *Center for Molecular and Cellular Biology, Skolkovo Institute of Science and Technology, Moscow, Russia*

[3] *Lomonosov Moscow State University, Moscow, Russia*

[4] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

** s.baltin@rambler.ru*

Key words: small open reading frame; sORF; uORF; mRNA; translation; *MCRS1*

Motivation and Aim: Small open reading frames (sORFs) are characterized by a short length (below 100 codons) compared to main open reading frames (mORF) encoding functional proteins. sORFs were considered nonfunctional initially, but later studies showed that many of them are involved in important physiological processes. sORFs are located in many types of RNA, both coding (mRNA) and non-coding, but most currently known sORFs are located in the 5' untranslated regions (5' UTR) of coding transcripts (upstream ORFs, uORF). Moreover, upstream ORFs may encode functional microproteins or play regulatory roles [1]. In particular, they may repress the translation of the downstream main ORFs or contribute to their translation in particular conditions [2]. The CRISPR-Cas9 screening previously performed in the laboratory has discovered that the disruption of uORFs in several human genes affects viability of near haploid HAP1 human cell line. One of the identified hits is located in *MCRS1* gene. Subsequent analysis showed that the *MCRS1* gene mRNA contains three small open reading frames in the 5' UTR, varying in length (sORF#1, sORF#2, sORF#3). The aim of the research is to validate sORFs translation and to study their biological function.
Methods and Algorithms: We obtained reporter constructions with the *MCRS1* 5' UTR to validate the translation of its sORFs and to find out the role of the longest sORF#1 in mORF translation. We have done genome editing with CRISPR-Cas9 for endogenous tagging of the sORF with HiBiT-tag to prove the existence of the micropeptide *in vivo*.
Results: We showed that the sORF#1 of the *MCRS1* gene is translated. sORF#1 encodes a 46 amino acid micropeptide, which is conserved among mammals. The presence of this micropeptide was proven *in vivo* using endogenous labeling and subsequent detection of the labeled micropeptide. Mutations in *MCRS1* sORF#1 introduced by CRISPR-Cas9 decrease cell viability. Furthermore, *MCRS1* sORF#1 represses the translation of the main ORF. The overexpression of the 5' UTR of *MCRS1* gene with HiBiT-tag insertion demonstrated that the shorter sORF#2 can also be translated. Using the dual luciferase reporter with cloned 5' UTR of the *MCRS1* gene, we found out that the second start codon of the mORF is involved in translation initiation; *MCRS1* translation predominantly begins with it. The first start codon of mORF takes almost no part in the initiation of *MCRS1* translation. Elongation of the sORF#1 does not affect the efficiency of translation of the mORF. sORF#1 shortening activates translation of the mORF. At the same time, we demonstrated that the *MCRS1* mRNA is not a subject to

regulation through nonsense-mediated decay as shown by the translation arrest using cycloheximide.

Conclusion: In conclusion, our study shows that sORF#1 regulates the translation of mORF of the *MCRS1* gene, and sORF#1 functioning is necessary for the cell viability. However, the detailed mechanism of the entire process remains unclear. Further research will elucidate the mechanisms of translation regulation by *MCRS1* sORF#1 and investigate the functions of the encoded microprotein to provide a complete picture of sORF functioning and open up new prospects for further research.

Funding: The study was supported by the Russian Science Foundation (23-14-00058).

References

1. Kute P.M., Soukarieh O., Tjeldnes H., Trégouët D.A., Valen E. Small Open Reading Frames, How to Find Them and Determine Their Function. *Front Genet*. 2022;12:796060. doi 10.3389/fgene.2021.796060
2. Kurihara Y. uORF Shuffling Fine-Tunes Gene Expression at a Deep Level of the Process. *Plants (Basel)*. 2020;9(5):608. doi 10.3390/plants9050608

doi 10.18699/bgrs2024-11.2-06

Pathways for selecting ribosomal proteins for packaging into exosomes

Belova E.D.[1]*, Tamkovich S.N.[2], Shefer A.A.[3], Graifer D.M.[4]

[1] *Laboratory of Ribosome Structure and Function, Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[2] *Laboratory of Molecular Medicine, Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[3] *Laboratory of Molecular Medicine, Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[4] *Laboratory of Ribosome Structure and Function, Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

** e.karbolina@g.nsu.ru*

Key words: ribosomal proteins; exosomes; endosomal sorting complexes; breast cancer

Motivation and Aim: Many ribosomal proteins are “moonlighting” ones having various functions beyond their canonical role as constituents of the ribosome (extra-ribosomal functions). These functions often relate to human pathologies, primarily, to carcinogenesis, via their participation in various signaling pathways, e. g., in the p53-MDM2 pathway, which regulates keystone tumor suppressor p53. Besides, expression of many ribosomal proteins is considerably changed (upregulated or downregulated) in malignant cells. On the other hand, exosomes derived from both malignant and normal cells often contain specific sets of ribosomal proteins. These extracellular vesicles are generally considered to be responsible for the intercellular communication by transporting specific molecular cargo from the donor cell to the recipient one. Thus, one can assume that oncogenic (or onco-suppressor) properties of ribosomal proteins could be, at least in part, transferred from one cell to another by exosomes. In the case of ribosomal protein uS3 such kind transfer has been documented experimentally. In particular, it has been demonstrated that transport of uS3 by exosomes from the gastric cancer cells resistant to an antitumor drug cisplatin to the non-resistant ones makes the recipient cells cisplatin-resistant [1]. However, mechanisms of sorting of ribosomal proteins to exosomes remain unknown. One can suggest that this process is mediated by proteins of the endosomal sorting complex ESCRT, namely the HGS and/or Alix proteins, which trigger a cascade of reactions for the formation of exosomes and the packaging of various biopolymers in them [2]. The purpose of this work was identification of partners of ribosomal proteins from the ESCRT complexes, which might be the most likely participants in their packaging.

Methods and Algorithms: The work was carried out with the triple-negative human breast cancer cell line BT-549. In this study, we used and optimized the co-immunoprecipitation technique to better detect interactions between the human ribosomal protein uS3 and transport proteins HGS and Alix. Binding partners of uS3 were identified with the use of Western blotting with the use of specific antibodies against proteins under investigation/ Besides, we used recombinant ribosomal protein uS3 to simulate cellular ribosomal stress conditions, upon which the level of free ribosomal proteins considerably increase.

Results: We did not find interactions between uS3 and HGS or Alix proteins at normal physiological conditions. Moreover, such interactions were not detected even in the presence of cross-linking agent formaldehyde that fix protein-protein complexes. On the other hand, we clearly demonstrated co-immunoprecipitation of uS3 with HGS or Alix proteins in cell lysates supplied with recombinant His-tagged uS3. Obviously, a low level of ribosome-free uS3 in cells under normal conditions does not allow detection of its interaction with transporting proteins. However, when concentration of uS3 considerably increases (which imitate cellular ribosomal stress, when ribosomal assembly is impaired), the interactions become well detectable. Under these conditions, we indeed demonstrated the ability of HGS and Alix proteins to bind uS3 in lysates of breast cancer cells.
Conclusion: The results obtained indicate that ESCRT components – proteins HGS and Alix can interact with ribosomal protein uS3 when the latter is present at a concentration considerably higher than that in normal breast cancer cells. These results support the suggestion on the role of these proteins in sorting of uS3 to exosomes, but they can play this role only under conditions of abnormally high level of free ribosomal protein in the cell, which could take place, for example, under stress conditions when ribosomal subunits assembly is impaired. Several questions remain unresolved: is there an interaction between other ribosomal proteins and the transporting proteins of the ESCRT complex? Is this kind interaction typical for other cell malignant lines and for non-malignant cells? Finally, it would be essential to understand whether intercellular transfer of ribosomal proteins by exosomes contribute to cancer progression and metastasis. All this is the next frontier of investigations.
Funding: The study is supported by the Russian Science Foundation (grant No. 23-24-00159).

References

1. Sun M. et al. Cisplatin-resistant gastric cancer cells promote the chemoresistance of cisplatin-sensitive cells via exosomal RPS3 mediated PI3K-Akt-cofilin-1 signaling axis. *Front Cell Dev Biol*. 2021;9:142-159. doi 10.3389/fcell.2021.618899
2. Henne W.M., Buchkovich N.J., Emr S.D. The ESCRT pathway. *Dev Cell*. 2011;19(1):77-91. doi 10.1016/j.devcel.2011.05.015

doi 10.18699/bgrs2024-11.2-07

Hotspots in ribosomal proteins – determinants of the translation machinery and beyond

Graifer D.*, Malygin A., Bulygin K., Babaylova E., Ochkasova A.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** graifer@niboch.nsc.ru*

Key words: ribosomal protein; eukaryotes; human cells; translation; amino acid residues; functional assignment

Motivation and Aim: Ribosome is a giant ribonucleoprotein machinery, which synthesizes proteins in all organisms from bacteria to human according to the program incoming as messenger RNA (mRNA) copied from definite parts of the genomic DNA. It is composed of two subunits; each consists of ribosomal RNAs and several dozens of ribosomal proteins. The structures of ribosomes from all kingdoms of life to date are deciphered at near-atomic level thanks to achievements in X-ray crystallography and cryo-electron microscopy. This structural information gives the rationale for the obtained earlier data on the arrangements of ribosomal ligands (mRNAs and tRNAs) in the respective functional sites of the mammalian ribosome obtained by site-directed cross-linking with the use of chemically or photochemically reactive tRNA and mRNA derivatives [1]. Specific amino acid residues of ribosomal proteins (RPs) located very close (within 3–5 angstroms) to mRNAs and tRNAs at these sites have been determined. It is assumed that these residues make contacts with the respective ligand, but structural information itself does not provide an idea on contributions of these contacts to the translation process and its regulation. In particular, it remained unclear which consequences would take place upon changes at amino acid residues contacting mRNAs and tRNAs. This information has been obtained recently in our studies described below.
Methods and Algorithms: To study contribution of specific amino acid residues of RPs to the translation process, we carried out investigations with human HEK293T cells, in which target RPs contained substitutions or deletions of amino acid residues that had been found earlier in close proximity to ribosomal ligands. Cells were transfected with plasmid constructions encoding FLAG-tagged wild type or mutant RP under investigation, and lysate of these cells were centrifuged in sucrose density gradients to analyze the content of the labeled RP in fractions of ribosomal subunits, 80S ribosomes and polysomes. If the mutation interferes the subunits assembly, the level of mutant protein is lower than that of the wild type RP in all fractions. If it stops translation at the initiation step, mutation does not affect the content of the RP in fractions of the subunits but it decreases in fractions of 80S ribosomes and polysomes. Finally, if the mutation blocks elongation, labeled RP content decreases only in fractions of polysomes. Then experiments *in vitro* enables identification the particular step of initiation or elongation affected by the mutation.
Results: We studied functional roles in translation of those amino acid residues of RPs, which had been identified earlier to cross-link to tRNA and mRNA derivatives [1]. In particular, the C-terminal portion of the RP uS19 (RPS15) (amino acid residues in positions 130–145) was found in the decoding site of the mammalian 40S ribosomal subunit; it has no homology in bacterial counterparts. The application of HEK293T cells

producing mutant RP uS19 lacking the C-tail and analysis of the ribosomal material from these cells have revealed that the tail is not important for the assembly of the post-initiation 80S ribosome but is essential for the correct accommodation of incoming aminoacyl-tRNA (aa-tRNA) at the A site [2]. This step of the elongation cycle is an essential prerequisite enabling the participation of the cognate aa-tRNA in the peptide bond formation. RP uS3 (RPS3) is exposed at the solvent side of the 40S subunit near the mRNA entry pore and its tetrapeptide 60-GEKG-63 possesses an ability, unique among RPs, to selectively cross-link to single-stranded model RNAs containing aldehyde groups, e. g., an abasic site [3]. We found that this tetrapeptide is critically important for maintaining the correct structure of the 48S preinitiation complex competent for the eIF5A-dependent binding of the 60S subunit, the way by which 80S ribosome is formed upon the completion of initiation of translation. Conserved motif YxxPKxYxK of the RP eS26 (RPS26) neighbors mRNA upstream of the E site bound codon. Replacement of the conserved residues of this motif for alanines did not significantly affect the bulk translation but led to accumulation of the light polysomes containing initiation factor eIF3 that did not leave the ribosome at elongation. Detailed analysis enabled the conclusion that this conserved RP eS26 motif is implicated in fine tune regulation of translation of selected mRNAs to maintain balance of proteins currently required for the cell life [4]. Finally, we determined functional role of the conserved motif 45-QSGYGGQTK-53 in the RP eL42 (L36AL) located close to the 3'-terminus of deacylated tRNA at the P site after peptide bond formation (in this state, the acceptor stem of the tRNA is shifted to the E site, which is known as "hybrid" P/E state). It turned out that the eL42 motif plays a critical role in the ability of the human ribosome to perform properly elongation cycle at the step of deacylated tRNA dissociation from the E site in the human cell [5].

Conclusion: The application of the approach based on the use of human cells producing labeled target mutant RP with alterations in the region of interest and the subsequent analysis of the incorporation of the mutant RP into ribosomal fractions allowed us to reveal for the first time translational assignment of specific RPs amino acid residues contacting tRNA and mRNA. Some of these assignments turned out to be not those that were expected from the structural data.

Funding: The study is supported by the Russian Foundation for Basic Research (grant 20-04-00400 to D.G.) and by the Russian State funded budget project of ICBFM SB RAS (reg. No. 121031300041-4).

References

1. Graifer D.M., Karpova G.G. Roles of ribosomal proteins in the functioning of translational machinery of eukaryotes. *Biochimie.* 2015;109:1-17
2. Bulygin K., Malygin A., Gopanenko A., Graifer D., Karpova G. The functional role of the C-terminal tail of the human ribosomal protein uS19. *Biochim Biophys Acta Gene Regul Mech*. 2020;1863:194490
3. Babaylova E., Malygin A., Gopanenko A., Graifer D., Karpova G. Tetrapeptide 60–63 of human ribosomal protein uS3 is crucial for translation initiation. *Biochim Biophys Acta Gene Regul Mech*. 2019;1862:194411
4. Bulygin K.N., Malygin A.A., Graifer D.M., Karpova G.G. The functional role of the eukaryote-specific motif YxxPKxYxK of the human ribosomal protein eS26 in translation. *Biochim Biophys Acta Gene Regul Mech.* 2022;1865:194842
5. Bulygin K.N., Malygin A.A., Graifer D.M. Functional involvement of a conserved motif in the middle region of the human ribosomal protein eL42 in translation. *Biochimie.* 2024;218:96-104

doi 10.18699/bgrs2024-11.2-08

Integration of bioinformatic tools for assessing promoter activity and gene translation elongation efficiency to predict microbial gene expression levels

Klimenko A.I.[1, 2]*, Fairushina K.S.[3], Ponomarenko M.P.[1], Kolchanov N.A.[1, 2], Lashin S.A.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *klimenko@bionet.nsc.ru*

Key words: translation efficiency; protein abundance prediction; codon usage

Motivation and Aim: Prediction of gene expression in prokaryotes is a fundamental bioinformatics problem that is significant for further advances in both basic research and biotechnology. The gene expression values, explicated as the amount of active protein, depend on various factors, primarily transcriptional and translational regulation. Therefore, it is important to evaluate and integrate these upstream data to accurately predict gene expression. Contextual and structural characteristics determining promoter activity have a serious impact on the transcriptional activity of the corresponding genes, while in some microorganisms it has been shown that translation elongation is the second most important factor determining the amount of protein in the cell after the mRNA level. In this regard integration of bioinformatic tools for assessing promoter activity and gene translation elongation efficiency is relevant for solving the problem of predicting microbial gene expression levels.

Methods and Algorithms: The solution to this problem requires protein-level expression profiles covering a significant portion of the microorganism's genes. Within the scope of this study, we used the data gathered for prokaryotes representing various phylogenetic groups from Pax-DB database, which contains protein expression data with sufficiently high gene coverage [1]. For the provided genomes the translation efficiency features for each protein-coding gene have been calculated using the EloE tool [2]. The ACTIVITY [3] system implements an approach to predicting the activity of a site based on its primary nucleotide sequence, in particular the activity of promoters, on the basis of the PROPERTY database containing information on the physical and chemical properties of DNA. Regression models of gene expression at the protein level incorporating bioinformatic characteristics of gene translation elongation efficiency and corresponding promoter activity were built using Python tools.

Results and Conclusion: In this work, we have been integrating software tools aimed at predicting the efficiency of various stages of gene expression, namely translation elongation and promoter activity of the corresponding genes. The designed framework allows us to expand the set of considered predictors of expression level, including both context-dependent conformational and physico-chemical properties of associated sequences and various derived indices evaluating the properties associated with the efficiency of translation elongation. The integration of these approaches with

experimental expression data obtained at different levels appears promising for future gene expression prediction studies.

Funding: The study is supported by the Budget Project FWNR-2022-0020.

References

1. Huang Q., Szklarczyk D., Wang M., Simonovic M., Mering C. von PaxDb 5.0: Curated Protein Quantification Data Suggests Adaptive Proteome Changes in Yeasts. *Mol Cell Proteomics*. 2023;22(10):100640. doi 10.1016/j.mcpro.2023.100640
2. Sokolov V., Zuraev B., Lashin S., Matushkin Y. Web application for automatic prediction of gene translation elongation efficiency. *J Integr Bioinform.* 2015;12(1):256. doi 10.2390/biecoll-jib-2015-256
3. Ponomarenko M.P., Ponomarenko J.V., Frolov A.S., Podkolodny N.L., Savinkova L.K., Kolchanov N.A., Overton G.C. Identification of sequence-dependent DNA features correlating to activity of DNA sites interacting with proteins. *Bioinformatics*. 1999;15(7-8):687-703. doi 10.1093/bioinformatics/15.7.687

doi 10.18699/bgrs2024-11.2-09

ADP-ribosylation of human ribosomal proteins by enzymes PARP1 and PARP2 *in vitro*

Krasnikov A.*, Naumenko K., Graifer D., Malygin A., Lavrik O.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** aka2014s@mail.ru*

Key words: ADP-ribosylation; HPF1; ribosomal protein; PARP1/2

Motivation and Aim: ADP-ribosylation is a posttranslational modification of proteins that plays a key role in many regulatory mechanisms and cellular processes. The modification is performed by enzymes of the PARP family (ARTD) and results in the attachment of one (mono-ADP-ribosylation) or several (poly(ADP-ribosyl)ation) of ADP-ribose units to proteins [1]. Previously, it was believed that ADP-ribosylation occurs mainly during DNA repair, but recently a lot of data has been obtained on the participation of PARP enzymes in a variety of cellular processes, in particular, in the regulation of ribosome biogenesis and gene expression at the translation level [2]. Proteomic analysis revealed many ADP-ribosylated proteins, including several dozen ribosomal proteins (RPs). However, it is still unknown which enzymes of the PARP family under what conditions modify certain RPs and what consequences their ADP ribosylation leads to in the cell. The study of this type of modification is an urgent task, since violations of the assembly or operation of ribosomes, as well as many of the non-ribosomal functions performed by RPs, are associated with the development of a number of neurodegenerative diseases and are associated with carcinogenesis.
Methods and Algorithms: Modification of human ribosomal proteins with PARP enzymes was carried out *in vitro* by incubation of reaction mixtures containing the enzyme PARP1/2, total ribosomal protein from 40S or 60S ribosomal subunits from human placenta, DNase I-treated DNA as an enzyme activator, co-factor HPF1, and [^{32}P]-NAD^{+}. The labeled protein products of the ADP-ribosylation reaction were separated in an SDS-containing polyacrylamide gel followed by radioautography.
Results: We show that ribosomal proteins of the human large (60S) ribosomal subunit of are able to effectively undergo ADP-ribosylation by enzymes PARP1 and PARP2. The dependence of the modification reaction on the presence of the HPF1 cofactor, DNA activator, as well as on the amount of substrate (total 60S subunit ribosomal protein) was studied. It was found that modification occurs only in the simultaneous presence of the DNA activator and the HPF1 cofactor. Almost all the same patterns with 60S RPs are observed with PARP2. An unexpectable result was that when the concentration of the substrate (RPs) increases to a certain threshold, the modification reaction is completely inhibited (both modification of RPs and automodification of the PARP). Pre-incubation of RPs with proteinase K restores automodification of PARP, which indicates that the effect of complete inhibition of the reaction is caused by the intact ribosomal proteins. The identification of ribosomal proteins modified by each of the PARPs is the task of further research. Remarkably, under conditions used for modification of ribosome-free RPs, the mature ribosomal subunits were not modified by the enzyme PARP1.
Conclusion: For the first time, we found out which enzymes and under what conditions ADP-ribosylate human ribosomal proteins. It was shown for the first time that the

reaction depends on the simultaneous presence of a DNA activator and a HPF1 cofactor, and is completely inhibited when the concentration of ribosomal proteins reaches a certain threshold value.

Funding: The study is supported by the Russian State funded budget project of ICBFM SB RAS (No. 121031300041-4).

References

1. Alemasova E.E., Lavrik O.I. Poly(ADP-ribosyl)ation by PARP1: reaction mechanism and regulatory proteins. *Nucleic Acids Res*. 2019;47(8):3811-3827
2. Kim D.S. et al. PARPs and ADP-ribosylation in RNA biology: from RNA expression and processing to protein translation and proteostasis. *Genes Dev*. 2020;34(5-6):302-320

doi 10.18699/bgrs2024-11.2-10

Two-domain HMGB proteins HMGB1 and HMO1 have different effects on the structure of nucleosomes and chromatosomes

Malinina D.[1]*, Sivkina A.[1], Khodyreva S.[2], Lavrik O.[2], Studitsky V.[1, 3], Feofanov A.[1, 4]

[1] *Biology Faculty, Lomonosov Moscow State University, Moscow, Russia*

[2] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

[3] *Fox Chase Cancer Center, Philadelphia, USA*

[4] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

* *daria99malinina@gmail.com*

Key words: nucleosome; high mobility group B; HMO1; HMGB1; H1 histone; spFRET

Motivation and Aim: HMGB (high mobility group B) is an abundant family of architectural non-histone proteins characterized by the presence of DNA-binding domain called HMG-box, which also has DNA bending activity. In cells these proteins are actively involved in DNA repair, transcription, replication and recombination. Homologous proteins HMO1 (yeast) and HMGB1 (mammalian) are members of this family. Both proteins contain two DNA binding boxes A and B, as well as a disordered C-terminal region. It is assumed that, despite their structural similarity, the roles of these proteins in the organization of chromatin are different: HMO1 stabilizes chromatin by acting as a linker histone [1]; HMGB1 promotes destabilization of the nucleosome structure and thereby facilitates the binding of chromatin remodelers and transcription factors to the nucleosomal DNA [2]. However, these hypotheses about the effect of HMO1 and HMGB1 on the structure of nucleosomes and chromatosomes need to be experimentally evaluated.

Methods and Algorithms: Mononucleosomes were assembled on fluorescently labeled DNA containing the 603 Widom nucleosome positioning sequence. Donor (Cy3) and acceptor (Cy5) fluorophores were situated at neighboring gyres of nucleosomal DNA or at linkers. Chromatosomes were formed by incubating nucleosomes with linker histone H1.0. Formation of the complex between HMGB protein (HMO1/HMGB1) and nucleosomes was studied by the electrophoretic mobility shift assay. Structural changes in nucleosomes and chromatosomes were investigated using single particle Förster resonance energy transfer (spFRET) microscopy [3].

Results: It is found that several HMO1 molecules bind to a nucleosome, affect conformation of nucleosomal DNA and increase a distance between the helices of linker DNA [4]. It was revealed that HMO1 counteracts structural changes in the region of linker DNA caused by the linker histone H1.0 during the chromatosome formation. At the same time, H1.0 and HMO1 appear to bind to independent sites in the nucleosome, so H1.0 remains bound in the presence of bound HMO1 [4].

HMGB1 is found to bind to the nucleosome in a 1:1 stoichiometry, and the mode of binding is independent of the presence or absence of linker DNA. HMGB1 binding does not affect the conformation of nucleosomal DNA, but decreases the distance between the linker DNA helices. HMGB1 partially destabilizes chromatosomes by releasing linker DNA from the complex with H1.0 histone.

Conclusion: HMO1 and HMGB1 differ significantly in their effect on the structure of nucleosomes and chromatosomes that can be a reason of their differences in the functional activity *in vivo*.

Funding: This study was supported by the Russian Science Foundation (No. 19-74-30003).

References

1. Panday A., Grove A. The high mobility group protein HMO1 functions as a linker histone in yeast. *Epigenet Chromatin*. 2016;9:13. doi 10.1186/s13072-016-0062-8
2. Travers A.A. Priming the nucleosome: a role for HMGB proteins? *EMBO Rep*. 2003;4(2):131-136. doi 10.1038/sj.embor.embor741
3. Kudryashova K.S., Chertkov O.V., Nikitin D.V. et al. Preparation of mononucleosomal templates for analysis of transcription with RNA polymerase using spFRET. *Methods Mol Biol*. 2015;1288:395-412. doi 10.1007/978-1-4939-2474-5_23
4. Malinina D.K., Sivkina A.L., Korovina A.N. et al. Hmo1 Protein Affects the Nucleosome Structure and Supports the Nucleosome Reorganization Activity of Yeast FACT. *Cells*. 2022;11(19):2931. doi 10.3390/cells11192931

doi 10.18699/bgrs2024-11.2-11

Numeric structure of genetic code in natural evolution: energy insight

Moldavanov A.
2774 Sunnybridge Drive, Burnaby, Canada
trandrei8@gmail.com

Key words: energy evolution; preferable nucleotides; genetic code; codon; triplet; quadruplet

Motivation and Aim: Earlier was shown that the process of energy evolution in open thermodynamic system (OTS) is characterized by occurrence of the k dissimilar evolutional scenarios [1]. On other hand, it is well known the critical role that the preferable nucleotides such as adenine (A), cytosine (C), guanine (G) and thymine (T) with segregated energy signatures play in assembly of deoxyribonucleic acid (DNA). So, the aim is to find the mechanism to reconcile coexistence of preferable nucleotides with structuring of energy evolution and support the known numeric appearance 64 (20) for total number of possible unique combinations in genetic code (GC).
Methods and Algorithms: This research is based on found analytic solution for interplay between two different stochastic processes. The first process is the flow of random quadruplets of DNA nucleotides A, C, G, T while the second one is permanent reproduction of structuring in process of energy exchange in DNA.
Results: It was shown that process of energy development in OTS features the discrete spectrum, where the special role belongs to first three harmonics that come as points of bifurcation (*PoB*) [2]. So, from an energy standpoint, process of exchange in DNA's assembly can be described from two angles. On one hand, DNA's assembly comes as a random codon-based realization of flow of 4 preferable nucleotides. On other hand, effective evolution is only possible if taken energy phase resides within the 3-part bifurcation area. Hence, we have to find a compromising solution for coexistence of 4 nucleotides within 3 acceptable *PoB* areas. As a result, in combinators terms we come to a classic problem for placement of 4 "balls" (nucleotides A, C, G, T) over 3 "bins" (*Phase 1, Phase 2, Phase 3*) as shown in the Figure. If order of "balls" is important, then solution for numerical structure of *GC* yields the total number of 64 unique codon combinations, otherwise of 20. So, we see that each evolutionary step is accompanied by realization 1 out of 64 (20) unique combinations, which squeezes available chemicals A, C, G, T into the *PoB* triplet (Fig. 1).

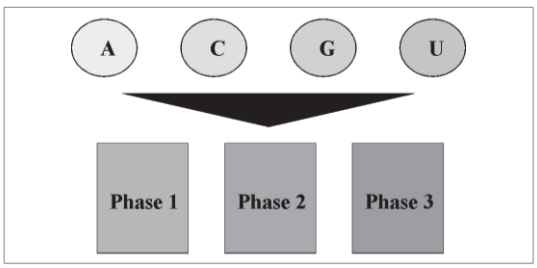


Fig. 1. Schematic setup for draw in placement of 4 nucleotides (A, C, G, U) over 3 available energy phases of different evolutional context (*Phase 1, Phase 2, Phase 3*)

Conclusion: Above results give a key for understanding of the possible way for forming of the known numerical structure of GC. The 64 (20)-scenario matrix comes as the work-horse for practical realization of GC algorithm. Nature drives above matrix repeatedly, creating codon chains from the quite limited number of available elements.
Funding: This study is not supported by any business, government, or science organization.

References

1. Moldavanov A. energy infrastructure of evolution for system with infinite number of links with environment. *BioSystems.* 2022;213:104607. doi 1016/j.biosystems.2022.104607
2. Moldavanov A.V. Time translation symmetry breaking as factor of system energy evolution. *Eur Phys J Plus.* 2022;137:1322. doi 10.1140/epjp/s13360-022-03550-5

doi 10.18699/bgrs2024-11.2-12

Search for upstream open reading frames that are essential for viability of human cell lines

Shepelev N.M.[1, 2]*, Kiniry S.J.[3], Lavrov A.I.[1], Razumova E.A.[1], Suspitsyna A.D.[4], Baranov P.V.[3], Dontsova O.A.[1, 2, 4], Rubtsova M.P.[1, 2]

[1] *Lomonosov Moscow State University, Moscow, Russia*

[2] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

[3] *University College Cork, Cork, Ireland*

[4] *Skolkovo Institute of Science and Technology, Moscow, Russia*

* *nikita.shepelev96@gmail.com*

Key words: small ORFs; sORFs; upstream ORFs; uORFs; CRISPR-Cas9; microproteins

Motivation and Aim: The Human Genome Project concluded over 20 years ago. However, the complete repertoire of coding open reading frames (ORFs) remains uncertain. Among these, the identification of small open reading frames, those consisting of less than 100 codons, poses a significant challenge [1]. Ribosome profiling has made a revolution in our understanding of translated regions of different RNA molecules [2]. Among the plethora of small ORFs, particular attention is drawn to those located in the 5' untranslated regions of mRNAs, commonly known as upstream ORFs. Upstream ORFs can regulate translation *in cis*, encode small functional peptides (microproteins), be a source of antigenic peptides, or serve as the origin for new proteins and their domains during the evolutionary process. It is estimated that nearly half of mammalian mRNAs contain upstream ORFs. Nevertheless, their functions largely remain understudied [3]. Therefore, we performed a functional genomics screening to identify essential upstream ORFs in human cell lines.

Methods and Algorithms: We utilized ribosome profiling data to predict upstream ORFs in human mRNAs, ultimately identifying 977 highly conserved targets. We then designed highly specific guide RNAs to target these upstream ORFs employing the CRISPR-Cas9 system. Our investigation involved an extensive study of selected upstream ORFs using the powerful CRISPR-Cas9 knockout screening technique to assess their influence on cell viability of HAP1, A549, and HEK293T human cell lines. Selected targets were verified by a competition assay using individual guide RNAs. Translation of upstream ORFs was confirmed using a reporter construct or by endogenous tagging. Using a dual-luciferase reporter, we tested the effect of targets on the translation of the main downstream ORFs.

Results: CRISPR-Cas9 knockout screening allowed us to identify numerous upstream ORFs potentially essential for human cell viability. Subsequent functional analysis confirmed the translation of several upstream ORFs by endogenous tagging or exogenous tagging in the reporter construct. Notably, several upstream ORFs were found to play a regulatory role in the translation of the main downstream ORFs. Conservation analysis coupled with peptide motif search indicates that some upstream ORFs may encode functional microproteins.

Conclusion: The CRISPR-Cas9 knockout screening enabled us to identify upstream ORF that play important roles in essential cellular processes. The following analysis

indicates that these selected upstream ORFs have the capability to modulate the translation of main ORFs or potentially encode functional microproteins.
Funding: The study is supported by the Russian Science Foundation (23-14-00058).

References

1. Schlesinger D., Elsässer S.J. Revisiting sORFs: overcoming challenges to identify and characterize functional microproteins. *FEBS J*. 2022;289(1):53-74. doi 10.1111/febs.15769
2. Ingolia N.T., Brar G.A., Stern-Ginossar N. et al. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep*. 2014;8(5):1365-1379. doi 10.1016/j.celrep.2014.07.045
3. Renz P.F., Valdivia-Francia F., Sendoel A. Some like it translated: small ORFs in the 5'UTR. *Exp Cell Res*. 2020;396(1):112229. doi 10.1016/j.yexcr.2020.112229

doi 10.18699/bgrs2024-11.2-13

m6A-dependent regulation of Cap-dependent translation

Smolin E.*, Buyan A., Lyabin D., Kulakovskiy I., Eliseeva I.
Institute of Protein Research, RAS, Pushchino, Russia
** smolinegoralexeyevich@gmail.com*

Key words: Epitranscriptomics; METTL3; METTL14; mRNA modifications; m6A, mTOR, regulation of translation

Motivation and Aim: The role of m6A methylation in mRNA of the regulation of translation is currently well-covered. When METLL3 is knockdown (i. d. methylated level decrease), total protein synthesis becomes more sensitive to mTOR inhibitors. This implies that under normal conditions a significant part of mRNA is methylated, and under cap-dependent translation inhibition conditions (e. g., stress inhibiting mTOR kinase activity) it allows mRNA to be effectively translated. However, the issue of the mechanism for working and regulation of the process is still open to settle. According to the general assumption, the eIF3 factor is binding with m6A in the 5'-UTR mRNA, which results in the ribosome planting on mRNA without the cap-structure [1]. However, it is not known how the whole transcriptome show changes in response to the inhibition of the mTOR signal pathway in methylase-free cells (METTL3 or METTL14).
Methods and Algorithms: The HEK293T cell line was employed in this study. For this genes knockdowns, the transfection of siRNA using Lipofectamin was employed. The protein expression level was evaluated by the analysis of the Western-blot. HEK293T cells, together with HEK293T knockdown METTL3/METTL14, were cultured in the absence and presence of Torin1 for 2 hours, after which cells were collected for NGS library preparation [2].
Deletion of adapters, filtering of reads that were too short, and deduplication were performed using CutAdapt (v2.10) and seqkit (v0.10.1). Mapping and read counting were conducted using STAR (v2.7.6), the standard human genome assembly hg38, by means of GENCODE basic gene annotation (v.38). Gene filtering was applied based on the number of reads from RNA-Seq and Ribo-Seq, with the thresholds of 5 counts per million (cpm) for RNA-Seq (the minimum library size of 2.2 million reads) and 15 cpm for Ribo-Seq (the minimum library size of 0.7 million reads). The read normalization and differential translation were computed using DESeq2 (v1.38.1), with the normalization performed by means of PSP73; GFP spike-ins for RNA-Seq, and Ribo spike-ins for Ribo-Seq. The analysis of RNA m6A methylation utilized data from DART-Seq, m6A-REF-Seq, miCLIP, MINES, and eTAM-Seq, originally obtained from HEK293T cells [3, 4].
Results: Experiments were conducted on HEK293T cell lines with the knockdown of methylases METTL3/14 (hereafter referred to as KD) and on wild-type (WT) cells under normal conditions and upon mTOR inhibition. Differential translation analysis identified several sets of genes exhibiting varying sensitivity to mTOR signaling pathway inhibition. The largest group, subsequently termed as “Down”, comprised 3849 mRNAs, demonstrating decreased translation levels in both WT and KD cells. As it was anticipated, this gene group was enriched with mRNAs containing TOP motifs [5], which are the primary targets of mTOR. Additionally, a group of genes (termed as

Neutral) was identified, where translation remained unchanged upon mTOR inhibition in both cell lines. Of particular interest is the group of genes (termed as KD Down), where translation levels of corresponding mRNAs decreased exclusively in cells with the knockdowns of METTL3/14 methylases.

To determine the proportion of methylated transcripts in each of the studied groups, the processed data from quantitative mapping of m6A modification (eTAM-Seq) were utilized. The analysis revealed an increasing proportion of methylated transcripts in the sequence Down < KD Down < Neutral. Based on it, it can be hypothesized that the sensitivity of different mRNA groups to mTOR inhibition depends on the degree of mRNA methylation.

Conclusion: We have demonstrated that the knockdown proteins of the methyltransferase complex METTL3 and METTL14 in HEK293T cells leads to the reduction in overall translation levels when the mTOR signaling cascade is inhibited. Analysis of the distribution of m6A modification in mRNA showed that the cumulative methylation level increases in the sequence Down < KD Down < Neutral. Thus, the most sensitive mRNAs to mTOR inhibition (the Down group) have the lowest methylation level. The elevated methylation levels are observed in the KD Down and Neutral groups. In wild-type cells, translation of such mRNAs is resistant to mTOR inhibition. However, when the amount of methyltransferase in the cell is reduced, leading to decreased mRNA methylation, translation is inhibited only in the KD Down group. It can be hypothesized that there is a certain threshold level of the mRNA methylation necessary for translation stability against mTOR inhibition. In wild-type cells, methylation below the threshold is a characteristic of the members of the Down group, whose translation is inhibited upon mTOR inhibition. With the reduced methyltransferase activity, the methylation level of the KD Down group members falls below the threshold, resulting in their translation being inhibited. Meanwhile, the members of the Neutral group, which a characterized by a higher methylation level, the reduction in methylation does not reach the threshold value, so their translation continues to be resistant to mTOR inhibition.

Funding: This study was supported by Russian Science Foundation (grant number 19-74-10079) and the Russian Foundation for Basic Research (grant number 20-34-90047).

References

1. Meyer K.D. et al. 5′ UTR m6A Promotes Cap-Independent Translation. *Cell*. 2015;163(4):999-1010
2. McGlincy N.J., Ingolia N.T. Transcriptome-wide measurement of translation by ribosome profiling. *Methods*. 2017;126:112-129
3. Capitanchik C. et al. How Do You Identify m(6) A Methylation in Transcriptomes at High Resolution? A Comparison of Recent Datasets. *Front Genet Switzerland*. 2020;11:398
4. Xiao Y.-L. et al. Transcriptome-wide profiling and quantification of N6- methyladenosine by enzyme-assisted adenosine deamination. *Nat Biotechnol*. 2023;41(7):993-1003
5. Anisimova A.S. et al. Multifaceted deregulation of gene expression and protein synthesis with age. *Proc Natl Acad Sci USA*. 2020;117(27):15581-15590

11

Симпозиум «Фундаментальные генетические/клеточные системы/процессы: вычислительные и экспериментальные подходы»

Symposium “Fundamental genetic/cellular systems/processes: computational and experimental approaches”

doi 10.18699/bgrs2024-11.3-01

Фосфорамидные бензимидазольные олигонуклеотиды в составе несовершенных комплексов с ДНК: гибридизационные свойства и эффективность удлинения в ходе ПЦР

Голышев В.[1]*, Морозова Ф.[1, 2], Бердюгин А.[1, 2], Ломзов А.[1]

[1] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *golyshevvictor@gmail.com*

Ключевые слова: аналоги нуклеиновых кислот; гибридизационные свойства; физико-химические свойства; однонуклеотидные несоответствия; ПЦР; праймеры; дуплексы

Мотивация и цель: С ростом требований к высокочувствительным и специфичным методам молекулярного обнаружения, совершенствовались методы диагностики, что привело к последовательному развитию эффективной техники детекции нуклеиновых кислот – полимеразной цепной реакции (ПЦР). Для повышения специфичности и чувствительности в составе праймеров для ПЦР используют различные типы модификаций нуклеотидов Одним из перспективных аналогов НК для этого могут выступать фосфорамидные азольные олигонуклеотиды (ФАО), метод автоматического твердофазного фосфитамидного синтеза которых был впервые недавно предложен в ИХБФМ СО РАН [1]. Уже начаты исследования их физико-химических [2] и молекулярно-биологических [3] свойств. Целью данной работы являлось изучение физико-химических свойств комплексов ДНК с новым классом модифицированных по межнуклеотидному фосфатному остатку дезоксирибонуклеиновых кислот – фосфорамидных бензимидазольных олигонуклеотидов, в которых будут присутствовать однонуклеотидные замены, а также исследование эффективности и специфичности удлинения модифицированных цепей, что послужит базисом для их дальнейшего применения.

Методы и алгоритмы: На первом шаге исследований использовали метод термической денатурации с оптической регистрацией сигнала в стандартных условиях (1 M NaCl, 10 mM $(CH_3)_2AsO_2Na$, pH 7.2) и в условиях, приближенным к условиям ПЦР (100 мМ NaCl, 5 мМ $MgCl_2$, pH 7.2), в большинстве экспериментов использовали суммарную концентрацию олигонуклеотидов 10 мкМ. Для этого использовали спектрофотометр Agilent Cary 3500 с 8-мисекционным держателем с Пельтье элементом для температурного контроля. Для анализа структуры модельных нативных и модифицированных комплексов были записаны спектры кругового дихроизма (КД) на спектрополяриметре Jasco J-600. Для изучения субстратных свойств ФАО/ДНК комплексов при процессировании их ДНК-зависимой ДНК-полимеразой использовали термоциклер Eppendorf Mastercycle Classic со следующей процедурой для одного цикла: состоящего из двух стадий: отжига – 95 °C – 2 минуты, 30 °C – 1 минута и элонгации при 37 °C в течении 7 минут. Электрофоретический анализ продуктов производили в 15 %

денатурирующем ПААГ. Сканирование гелей проводили на VersaDoc VH 4000 (Bio-Rad, Hercules, CA, USA) использую канал для FAM.

Результаты: В рамках данного исследования были изучены физико-химические свойства тринадцатизвенных олигодезоксирибонуклеотидов, содержащие одну N-бензимидазольную модификацию в концевой или предконцевой межнуклеотидной фосфатной группе с 3'- конца, формирующих комплексы с 22-звенными ДНК цепями. Формируемые ими комплексы были комплементарными или содержали мисматч во втором положении с 3'- конца короткой цепи. Было показано, что процесс образования таких комплексов может быть описан в рамках приближения модели двух состояний. Введение бензимидазольной модификации значительно снижает термическую стабильность таких комплексов в стандартных условиях (1M NaCl), но в условиях, близких к физиологическим такого снижения практически не наблюдается. Введение N-бензимидазольной модификации в первый фосфатный остаток приводит к меньшей дестабилизации дуплексов, чем ее введение во 2 положение от 3'-конца короткой цепи.

Было показано, что структура комплексов нативных и модифицированных комплексов без мисматчей практически совпадают и соответствуют B-форме двойной спирали ДНК. Наличие однонуклеотидных несоответствий и модификаций не приводит к заметному нарушению спиральной формы комплекса.

На следующем этапе была исследована скорость и эффективность элонгации тринадцатизвенных ФАО праймеров в комплексе с 22-звенной ДНК-матрицей как для систем с мисматчами, так и комплементарных комплексов при процессировании их ДНК-зависимой ДНК-полимеразой. Была изучена кинетика удлинения праймера на основании чего выбраны оптимальные условия анализа. Анализ продуктов удлинения ФАО-праймеров *Taq* ДНК-полимеразой позволил определить важные для действия фермента положения ФА-групп в праймере замедляющие или препятствующие удлинению олигомера. В частности, модификация первого межнуклеотидного фосфатного остатка с 3'-конца праймера приводила к заметному снижению эффективности его удлинения, в то время как модификация второго фосфата в меньшей степени ингибировала процесс элонгации как для комплементарных комплексов, так и комплексов с мисматчем.

Выводы: Таким образом, изучены физико-химические и субстратные свойства ФАО/ДНК комплексов (как имеющих в предконцевом положении однонуклеотидные несоответствия, так и комплементарные системы) при процессировании их ДНК-зависимой ДНК-полимеразой. Показано, что наличие бензимидазольной фосфорамидной модификации в первом межнуклеотидном фосфатном остатке от 3'-конца практически полностью ингибирует процесс элонгации в системах с однонуклеотидными несоответствиями.

Финансирование: Исследование поддержано грантом РНФ № 23-74-01116.

Phosphoramide benzimidazole oligonucleotides in imperfect complexes with DNA: hybridization properties and elongation efficiency during PCR

Golyshev V.[1]*, Morozova F.[1, 2], Berdugin A.[1, 2], Lomzov A.[1]

[1] *Institute of chemical biology and fundamental medecine, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *golyshevvictor@gmail.com*

Key words: nucleic acid analogs; hybridization properties; physicochemical properties; single nucleotide mismatches; PCR; primers; duplexes

Motivation and Aim: With the growing requirements for highly sensitive and specific methods of molecular detection, diagnostic methods have been improved, which led to the consistent development of an effective technique of nucleic acid detection – polymerase chain reaction (PCR). In order to increase specificity and sensitivity, various types of modifications of nucleotide derivatives are used as primers for PCR. One of the promising analogs of NA for this purpose can be phosphoramide azole oligonucleotides (PAO), the method of automated solid-phase phosphitamide synthesis of which was recently proposed for the first time in ICBFM SB RAS [1]. Studies on their physicochemical [2] and molecular biological [3] properties have already been initiated. The aim of this work was to study the physicochemical properties of DNA complexes with a new class of deoxyribonucleic acids – phosphoramide benzimidazole oligonucleotides modified on the inter-nucleotide phosphate residue, in which single-nucleotide substitutions will be present, as well as to study the efficiency and specificity of the modified chains elongation, which will serve as a basis for their further application.

Methods and Algorithms: In the first step of the research we used the method of thermal denaturation with optical signal registration under standard conditions (1 M NaCl, 10 mM $(CH_3)_2AsO_2Na$, pH 7.2) and under conditions close to PCR conditions (100 mM NaCl, 5 mM $MgCl_2$, 10 mM $(CH_3)_2AsO_2Na$, pH 7.2); in most experiments we used the total oligonucleotide concentration of 10 μM. For this purpose, an Agilent Cary 3500 spectrophotometer with an 8-section Peltier element for temperature control was used. To analyze the structure of model complexes of native and modified complexes, circular dichroism (CD) spectra were recorded on a Jasco 600 spectropolarimeter. To study the substrate properties of FAO/DNA complexes during their processing by DNA-dependent DNA polymerase, an Eppendorf Mastercycle Classic thermocycler was used with the following procedure for one cycle: consisting of two stages: annealing at 95 °C for 2 minutes, 30 °C for 1 minute and elongation at 37 °C for 7 minutes. Electrophoretic analysis of the products was performed in denaturing 15 % PAGE. Gels were scanned on a VersaDoc VH 4000 (Bio-Rad, Hercules, CA, USA) using the FAM channel.

Results: This study investigated the physicochemical properties of complexes of thirteen-stranded oligodeoxyribonucleotides containing a single N-benzimidazole modification in the terminal or preterminal phosphate group at the 3'-end with 22-stranded DNA strands. The complexes formed by them were complementary or contained mismatches in the second position from the 3'-end of the short chain. It was shown that the process of formation of such complexes can be described within the

approximation of the two-state model. The introduction of benzimidazole modification significantly reduces the thermal stability of such complexes under standard conditions (1M NaCl), but under conditions close to physiological conditions such reduction is practically not observed. The introduction of N-benzimidazole modification into the first phosphate residue leads to less destabilization than introduction into position 2 from the 3'-end of the short chain.

It was shown that the structure of the complexes of native and modified complexes without mismatches practically coincide and correspond to the B-form of the DNA double helix. The presence of single-nucleotide mismatches and modifications both 5' and 3' away from it does not lead to a noticeable disruption of the helical shape of the complex.

At the next stage, the rate and efficiency of elongation of thirteen-stranded FAO primers in complex with 22-stranded DNA primers were investigated for both the systems with mismatches and complementary complexes when they were processed by DNA-dependent DNA polymerase. The kinetics of primer elongation was studied and optimal conditions of analysis were chosen. The analysis of PAO-primer elongation products by Taq DNA polymerase allowed us to determine the positions of PA-groups in the primer that are important for the enzyme action and slow down or prevent oligomer elongation. In particular, modification of the first internucotide phosphate residue at the 3'-end of the primer led to a marked decrease in its elongation efficiency, while modification of the second phosphate inhibited the elongation process to a lesser extent for both complementary complexes and complexes with mismatch.

Conclusion: Thus, we have studied the physicochemical and substrate properties of PAO/DNA complexes (both those with single-nucleotide mismatches in the preterminal position and complementary systems) when they are processed by DNA-dependent DNA polymerase. It was shown that the presence of a benzimidazole phosphoramide modification in the first inter-nucleotide phosphate residue from 3' almost completely inhibits the elongation process in systems with single-nucleotide mismatches.

Funding: The study is supported by the Russian Science Foundation. (No. 23-74-01116).

Список литературы/References

1. Vasilyeva S.V. et al. Synthesis of oligonucleotides carrying inter-nucleotide N-(Benzoazole)-phosphoramide moieties. *ACS Omega*. 2022;8(1):1556-1566. doi 10.1021/acsomega.2c07083
2. GolyshevV.M. et al. Properties of phosphoramide benzoazole oligonucleotides (PABAOs). I. Structure and hybridization efficiency of N-benzimidazole derivatives. *Biochem Biophys Res Commun.* 2024;693:149390. doi 10.1016/j.bbrc.2023.149390
3. Chubarov A.S. et al. Phosphoramidate azole oligonucleotides for single nucleotide polymorphism detection by PCR. *Int J Mol Sci.* 2024;25(1):617. doi 10.3390/ijms25010617

doi 10.18699/bgrs2024-11.3-02

Изменения транскриптома в клетках глиомы при заражении онколитическим вирусом VV-GMCSF-Lact

Дымова М.А.[1]*, Семенов Д.В.[1], Васильева Н.С.[1], Кулигина Е.В.[1], Савиновская Ю.И.[1], Агеенко А.Б.[1], Мишинов С.В.[2], Рихтер В.А.[1]

[1] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

[2] *Новосибирский научно-исследовательский институт травматологии и ортопедии им. Я.Л. Цивьяна, Новосибирск, Россия*

** maya.a.rot@gmail.com*

Ключевые слова: глиома; клеточные культуры, полученные от пациентов; виротерапия; онколитический вирус; VV-GMCSF-Lact; транскриптом; дифференциально экспрессируемые гены

Мотивация и цель: Онколитическая виротерапия – это быстро развивающийся и перспективный метод, нацеленный на избирательное уничтожение опухолевых клеток. Ранее разработанный рекомбинантный вирус (VV-GMCSF-Lact) уже проходит первую фазу клинических испытаний в отношении рака молочной железы, однако он показал свою эффективность и против опухолевых клеток глиомы человека. Целью данной работы являлась оценка влияния вируса VV-GMCSF-Lact на иммортализованные и полученные от пациентов клетки глиомы человека на уровне их транскриптома.

Методы и алгоритмы: В работе были использованы иммортализованные и полученные от пациентов клетки глиомы человека. Для каждой культуры клеток была определена 50 % цитотоксическая доза (CD50) вируса VV-GMCSF-Lact. Данные клетки были инкубированы с вирусом VV-GMCSF-Lact в течение 12 и 24 часов. Далее был проведен транскриптомный анализ.

Результаты Анализ транскриптома в клетках через 12 или 24 часа после заражения вирусом VV-GMCSF-Lact показал общую активацию генов гистонов. Кроме того, повышенную экспрессию показали гены, связанные с ответом IFNγ, сигнальным путем NF-κB и воспалением. Напротив, гены, участвующие в регуляции клеточного цикла, включая организацию веретена деления, сегрегацию сестринских хроматид и контрольную точку G2/M, были подавлены после заражения вирусом. Гены, ответственные за транспорт белков и их секрецию, положительно коррелировали с CD50 вируса VV-GMCSF-Lact, а гены, кодирующие белки, участвующие в созревании ядерных транскриптов и сплайсинге, отрицательно коррелировали с CD50.

Выводы: В целом, результаты работы еще раз подчеркивают важность понимания молекулярных механизмов, сигнальных путей и генных сетей, влияющих на реакцию клеток глиомы на VV-GMCSF-Lact. Гены, экспрессия которых коррелирует с чувствительностью клеток к вирусу, представляют собой интерес в плане разработки таргетных препаратов, способных повысить эффективность виротерапии.

Финансирование: Поддержание клеточных культур выполнено в рамках государственного задания ИХБФМСО РАН №121030200173-6.

Transcriptome changes in glioma cells infected with the oncolytic virus VV-GMCSF-Lact

Dymova M.A.[1]*, Semenov D.V.[1], Vasileva N.S.[1], Kuligina E.V.[1], Savinovskaya Yu.I.[1], Ageenko A.B.[1], Mishinov S.V.[2], Richter V.A.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine, Siberian Branch, SB RAS, Novosibirsk, Russia*

[2] *Ya.L. Tsivyan Novosibirsk Research Institute of Traumatology and Orthopedics, Novosibirsk, Russia*

* *maya.a.rot@gmail.com*

Key words: glioma; patient-derived cell cultures; virotherapy; oncolyticvirus; VV-GMCSF-Lact; transcriptome; differentially expressed genes

Motivation and Aim: Oncolytic virotherapy is a rapidly developing and promising method aimed at selective destruction of tumor cells. A previously developed recombinant virus (VV-GMCSF-Lact) is already undergoing phase I clinical trials for breast cancer, but it has also shown its effectiveness against human glioma tumor cells. The aim of this work was to evaluate the effect of the VV-GMCSF-Lact virus on immortalized and patient-derived human glioma cells at the transcriptome level.

Methods and Algorithms: Immortalized and patient-derived human glioma cells were used in the study. The 50 % cytotoxic dose (CD50) of the VV-GMCSF-Lact virus was determined for each cell culture. These cells were incubated with the VV-GMCSF-Lact virus for 12 and 24 hours. Transcriptome analysis was then performed.

Results:Transcriptome analysis of cells 12 or 24 h after VV-GMCSF-Lact infection showed a general upregulation of histone genes. In addition, genes associated with the IFNγ response, NF-κB signaling pathway, and inflammation were upregulated. In contrast, genes involved in cell cycle regulation, including spindle organization, sister chromatid segregation, and the G2/M checkpoint, were downregulated after VV-GMCSF-Lact infection. Genes responsible for protein transport and secretion were positively correlated with VV-GMCSF-Lact CD50, while genes encoding proteins involved in the maturation of nuclear transcript and splicing were negatively correlated with CD50.

Conclusion: Overall, the results of the work once again highlight the importance of understanding the molecular mechanisms, signaling pathways and gene networks that influence the response of glioma cells to VV-GMCSF-Lact. Genes whose expression correlates with cell sensitivity to the virus are of interest in terms of developing targeted drugs that can improve the effectiveness of virotherapy.

Funding: Maintenance of cell cultures was supported by the Russian state-funded project for ICBFM SB RAS (grant number 121030200173-6).

doi 10.18699/bgrs2024-11.3-03

Влияние IGF-1 на апоптоз и число клеток в преимплантационных эмбрионах мышей, подвергнутых хроническому психосоциальному стрессу

Игонина Т.Н.[1]*, Лебедева Д.А.[1], Шавшаева Н.А.[1, 2], Амстиславский С.Я.[1]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *egik_00@mail.ru*

Ключевые слова: инсулиноподобный фактор роста I типа (IGF-1); хронический психосоциальный стресс; апоптоз; преимплантационные эмбрионы

Мотивация и цель: На сегодняшний день известно, что психосоциальный стресс может оказывать негативное влияние на женскую репродуктивную систему, однако механизмы, лежащие в основе этих эффектов, мало изучены. Фолликулогенез в яичниках регулируется как гонадотропинами гипофиза, так и внутриовариальными регуляторными факторами [1–4]. Одним из важных местных регуляторных факторов яичника является инсулиноподобный фактор роста 1 (IGF-1), он участвует на всех этапах фолликулогенеза, а именно на стадии рекрутирования фолликулов, отбора и доминирования [5, 6]. В предыдущих исследованиях мы показали, что хронический психосоциальный стресс ухудшает качество ооцитов у мышей, а также усиливает накопление активных форм кислорода и увеличивает процент апоптотических клеток в кумулюс-ооцитных комплексах. Кроме того, нами было показано снижение общего содержания IGF-1 в ткани яичника и снижение экспрессии генов *Igf-1* и *Igf-1r* в кумулюс-ооцитных комплексах при стрессе. В результате мы выдвинули гипотезу о том, что IGF-1 может участвовать в стресс-опосредованной реакции гипоталамо-гипофизарно-гонадной системы. Целью данной работы было исследование влияния введения IGF-1 *in vivo* на фоне хронического психосоциального стресса на общее число клеток и уровень апоптоза в эмбрионах мышей на стадии бластоцисты.

Методы и алгоритмы: Исследования проводили на мышах линии CD1. Животных содержали в стандартных условиях SPF-вивария ИЦиГ СО РАН. Все исследования были одобрены Комитетом по биоэтике ИЦиГ СО РАН (Протокол № 143 от 15.03.2023). Самки мышей в возрасте 8–12 недель были случайным образом распределены на две группы: Контроль (без стрессирующего воздействия) и Стресс (подвергнутые воздействию стрессирующих факторов). Хронический психосоциальный стресс был вызван следующим образом: самок мышей содержали в изоляции (по одной самке в клетке) в течение 11 дней, затем в последующие 10 дней помещали в условия скученности (по 11 самок в клетке). Животных контрольной группы содержали в стандартных условиях в группах по 5 самок на клетку. Эффективность применяемой процедуры для вызывания стрессовой реакции была подтверждена, поскольку уровень кортикостерона в сыворотке крови стрессированных мышей был повышен; определение проводили методом иммуноферментного анализа с использованием набора Elisa (Cloud-Clone Corp., США).

Для исследования получали преимплантационные эмбрионы от самок: 1) с естественным эстральным циклом, 2) после стимуляции яичников экзогенными гонадотропинами, 3) после введения IGF-1 на фоне введения гонадотропинов (группа Стресс) или 4) физиологического раствора на фоне введения гонадотропинов (группа Контроль). У самок мышей определяли стадию цикла по вагинальным мазкам, как описано ранее [7]. Для получения эмбрионов от мышей с естественным эстральным циклом самки на стадии эструса были ссажены с фертильными самцами. Для получения эмбрионов от мышей после стимуляции яичников самкам мышей на стадии диэструса вводили внутримышечно 10 МЕ гонадотропина сыворотки жеребых кобыл (ГСЖК, фоллигон, Intervet, Нидерланды), затем через 46 ч внутримышечно вводили 5 МЕ хорионического гонадотропина человека (ХГЧ, хорулон, Intervet, Нидерланды), что является общепринятым протоколом стимуляции яичников мышей [8]. После этого самки были ссажены с фертильными самцами. Для получения эмбрионов от мышей после введения IGF-1/физ. раствора, самкам мышей на стадии диэструса вводили внутримышечно IGF-1 (1 мкг в 0.1 мл физ. раствора)/физ. раствор (0.1 мл), через 48 ч повторно вводили IGF-1/физ. раствор и проводили гонадотропную стимуляцию яичников по вышеописанной схеме, затем ссаживали с фертильными самцами. День обнаружения вагинальной пробки считали первым днем беременности. На четвертый день беременности самок мышей подвергали эвтаназии с помощью CO_2, затем собирали эмбрионы для последующих исследований.

Исследование апоптоза в преимплантационных эмбрионах производили при помощи набора TUNEL (TdT-mediated dUTP nick end labeling) FITC Apoptosis Detection Kit (Vazyme, Китай), согласно протоколу производителя. Анализ проводили с помощью микроскопа AxioImager.M2 (Carl Zeiss, Германия) с фильтрами, подходящими для DAPI и FITC. Подсчет общего числа ядер и апоптотических клеток в эмбрионах осуществляли путем визуальной оценки препаратов с использованием программы ImageJ, как описано ранее [9].

Статистический анализ данных проводили с помощью двухфакторного дисперсионного анализа с последующим апостериорным сравнением с использование критерия Фишера LSD. Различия при значении $p < 0.05$ считали статистически значимыми.

Результаты: Анализ данных показал статистически значимое влияние хронического психосоциального стресса ($F(2, 437) = 9.3$, $p < 0.001$), а также введения гонадотропинов и IGF-1/физ. раствора ($F(4, 874) = 35.0$, $p < 0.001$) и взаимодействие этих факторов ($F(4, 874) = 7.6$, $p < 0.001$) на число клеток и уровень апоптоза в эмбрионах (см. рисунок). При естественном цикле на фоне хронического стресса наблюдалось сниженное число клеток ($p < 0.001$) и повышенное число апоптотических клеток ($p < 0.01$) в эмбрионах мышей по сравнению с контролем. Введение гонадотропинов приводило к снижению общего числа клеток в эмбрионах как в контрольной ($p < 0.001$), так и в стрессовой ($p < 0.001$) группах. При этом уровень апоптоза в контрольной группе оставался неизменным, тогда как у стрессированных самок на фоне гонадотропной стимуляции яичников происходило снижение процента апоптотических клеток в эмбрионах ($p < 0.01$). На фоне гонадотропной стимуляции яичников введение фактора роста IGF-1 самкам мышей, подвергнутых хроническому стрессу, приводило к увеличению числа клеток в эмбрионах ($p < 0.001$). Кроме того, после

введения IGF-1 у стрессированных самок общее число клеток ($p < 0.05$) и процент апоптотических клеток ($p < 0.05$) в эмбрионах были выше, чем у контрольных самок после гормональной стимуляции в сочетании с введением физ. раствора.

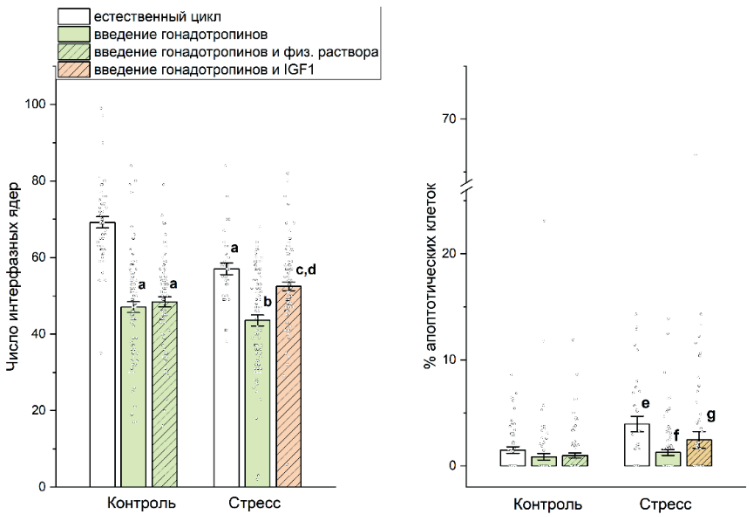


Рис. 1. Число клеток и уровень апоптоза в эмбрионах стрессированных и нестрессированных мышей. [a] $p < 0.001$, [e] $p < 0.01$ по сравнению с группой Контроль, естественный цикл; [b] $p < 0.001$, [f] $p < 0.01$ по сравнению с группой Стресс, естественный цикл; [c] $p < 0.001$ по сравнению с группой Стресс, стимуляция гонадотропинами; [d, g] $p < 0.05$ по сравнению с группой Контроль, стимуляция гонадотропинами и введение физиологического раствора

Выводы: Наше исследование показало, что хронический психосоциальный стресс ухудшает качество эмбрионов на стадии бластоцисты, выражающееся в сниженном числе клеток и повышенном уровне апоптоза. Стимуляция яичников гонадотропинами приводит к снижению числа клеток в эмбрионах как в контроле, так и при стрессе. При этом введение гонадотропинов на фоне стресса вызывало снижение процента апоптотических клеток в эмбрионах. Введение IGF-1 на фоне гонадотропной стимуляции яичников приводило к увеличению числа клеток в эмбрионах мышей, однако несколько повышало уровень апоптоза. В целом сочетание гонадотропной стимуляции с IGF-1 оказывает положительное влияние на качество эмбрионов мышей, подвергнутых психосоциальному стрессу.
Финансирование: Исследование поддержано Российским научным фондом (проект № 23-25-00139).

Effect of IGF-1 on apoptosis and cell number in preimplantation embryos of mice exposed to chronic psychosocial stress

Igonina T.N.[1]*, Lebedeva D.A.[1], Shavshaeva N.A.[1, 2], Amstislavsky S.Ya.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *egik_00@mail.ru*

Key words: insulin-like growth factor type I (IGF-1); chronic psychosocial stress; apoptosis; preimplantation embryos

Motivation and Aim: It is now known that psychosocial stress negatively affect the female reproductive system, but the mechanisms underlying these effects are poorly understood. Folliculogenesis in the ovaries is regulated by pituitary gonadotropins and intraovarian regulatory factors [1–4]. Insulin-like growth factor 1 (IGF-1) is one of the local regulatory factors important for all the stages of folliculogenesis, i. e. the recruitment, selection and dominance [5, 6]. In our previous studies, we have shown that chronic psychosocial stress causes low quality of oocytes in mice, as well as increases the accumulation of reactive oxygen species and elevates the percentage of apoptotic cells in cumulus-oocyte complexes. In addition, we demonstrated a decrease in the total IGF-1 content in ovarian tissue and a decrease in the expression of the *Igf-1* and *Igf-1r* genes in cumulus-oocyte complexes under stress condition. As a result, we hypothesized that IGF-1 may be involved in the stress-mediated response of the hypothalamic-pituitary-gonadal axis. The purpose of this work was to study the effect of IGF-1 *in vivo* administration under chronic psychosocial stress condition on the total number of cells and the level of apoptosis in mouse embryos at the stage of blastocyst.

Methods and Algorithms: The studies were carried out on CD1 mice. Animals were kept under standard conditions of SPF-Vivarium of the ICG SB RAS. All studies were approved by the Bioethics Committee of the Institute of Cytology and Genetics SB RAS (Protocol No. 143 of March 15, 2023). Female mice aged 8–12 weeks were randomly assigned to two groups: Control (no stressors) and Stress (exposed to stressors). Chronic psychosocial stress was induced as follows: female mice were kept in isolation (one female per cage) for 11 days, thereafter placed in overcrowded conditions (11 females per cage) for the next 10 days. Animals of the control group were kept under standard conditions in groups of five females per cage. The effectiveness of the procedure used to induce a stress response was confirmed by the enhanced level of corticosterone in the blood serum; measurement was done using an enzyme immunoassay using the Elisa kit (Cloud-Clone Corp., USA).

For the study, preimplantation embryos were obtained from females (1) with a natural estrous cycle, (2) after ovarian stimulation with exogenous gonadotropins, (3) after administration of IGF-1 alongside with the gonadotropin stimulation (Stress group) or (4) saline (Control group) alongside with the gonadotropin stimulation. In female mice, the cycle stage was determined by vaginal smears, as described previously [7]. To obtain embryos from mice within a natural estrous cycle, females at the estrus stage were mated with fertile males. To obtain embryos from mice after ovarian stimulation, female mice at the diestrus stage were injected intramuscularly with 10 IU of pregnant mares serum gonadotropin (PMSG, Folligon, Intervet, the Netherlands); 5 IU of human chorionic gonadotropin (hCG, Chorulon, Intervet, the Netherlands) was injected intramuscularly 46 h later, i. e. a standard protocol for ovarian stimulation in mice [8]. Thereafter, the females were mated with fertile males. To test effect of IGF-1, female mice at the diestrus stage were injected intramuscularly either with IGF-1 (1 μg of IGF-1 dissolved in 0.1 ml saline) or saline (0.1 ml); thereafter IGF-1/saline was repeated 48 h later, and standard ovarian stimulation with gonadotropins was applied as described above, as well as mating with fertile males. The day the vaginal plug was discovered was considered the first day of pregnancy. On the fourth day of pregnancy, female mice were euthanized

using CO_2, and then the embryos at the blastocyst stage were collected for the subsequent analysis.

Apoptosis in preimplantation embryos was studied using the TUNEL (TdT-mediated dUTP nick end labeling) FITC Apoptosis Detection Kit (Vazyme, China), according to the manufacturer's protocol. The analysis was carried out using an AxioImager.M2 microscope (Carl Zeiss, Germany) with filters suitable for DAPI and FITC. The total number of nuclei and apoptotic cells in embryos was counted by visual assessment using the ImageJ program, as described previously [9].

Statistical analysis of the data was performed using two-way analysis of variance followed by post hoc comparison using Fisher's LSD test. Differences with $p < 0.05$ were considered statistically significant.

Results: Data analysis showed a statistically significant effect of chronic psychosocial stress ($F(2, 437) = 9.3$, $p < 0.001$), as well as the administration of gonadotropins and IGF-1/saline ($F(4, 874) = 35.0$, $p < 0.001$) and the interaction of these factors ($F(4, 874) = 7.6$, $p < 0.001$) on the number of cells and the level of apoptosis in embryos (Fig. 1).

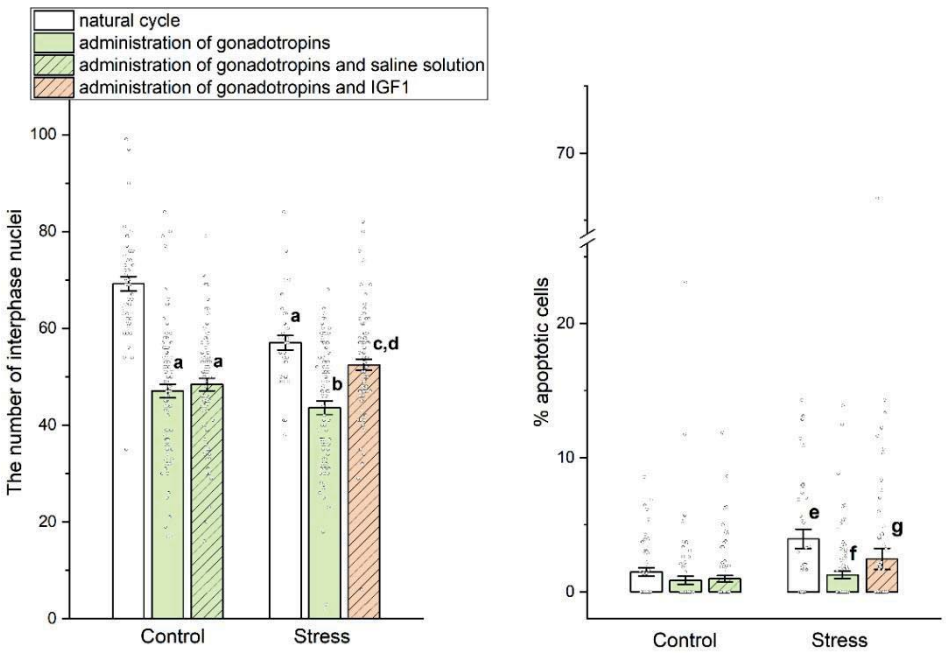


Fig. 1. Cell number and level of apoptosis in embryos of stressed and non-stressed mice.
[a] $p < 0.001$,[e] $p < 0.01$ compared to the Control group, natural cycle; [b] $p < 0.001$, [f] $p < 0.01$ compared with the Stress group, natural cycle; [c] $p < 0.001$ compared to the Stress group, stimulation with gonadotropins; [d, g] $p < 0.05$ compared with the Control group, stimulation with gonadotropins and administration of saline solution

During the natural cycle, in the conditions of chronic stress, a reduced total cell number ($p < 0.001$) and an increased number of apoptotic cells ($p < 0.01$) were observed in mouse embryos compared to the non-stressed controls. The administration of gonadotropins led to a decrease in the total number of cells in embryos of both non-stressed ($p < 0.001$) and stressed ($p < 0.001$) groups. The level of apoptosis in the control non-stressed group remained unchanged after treatment with gonadotropins, while in the stressed females, ovarian stimulation with gonadotropins caused a decrease of the percentage of apoptotic cells in the embryos ($p < 0.01$). Administration of the growth factor IGF-1 to female mice exposed to chronic stress alongside with gonadotropin treatment led to an increase in the number of cells in the embryos ($p < 0.001$). In addition, after the administration of IGF-1 to stressed females stimulated with gonadotropins, the total number of cells ($p < 0.05$) and the percentage of apoptotic cells ($p < 0.05$) in embryos were higher than in control females after hormonal stimulation in combination with saline administration.
Conclusion: Our study showed that chronic psychosocial stress in mice resulted in the low quality of embryos, indicated by the reduced number of cells and an increased level of apoptosis. Stimulation of the ovaries with gonadotropins leads to a decrease in the number of cells in embryos, both in non-stressed and stressed groups. Moreover, the administration of gonadotropins to stressed females caused a decrease in the percentage of apoptotic cells in embryos. Administration of IGF-1 together with ovarian stimulation with gonadotropins led to an increase in the number of cells in mouse embryos, but slightly increased the level of apoptosis. In general, the combination of gonadotropic stimulation with IGF-1 has a positive effect on the quality of embryos in mice exposed to psychosocial stress.
Funding: The study was supported by the Russian Science Foundation (project No. 23-25-00139).

Список литературы/References

1. di Clemente N., Racine C., Pierre A., Taieb J. Anti-Müllerian Hormone in Female Reproduction. *Endocr Rev*. 2021;42(6):753-782. doi 10.1210/endrev/bnab012
2. Chang C.W., Sung Y.W., Hsueh Y.W. et al. Growth hormone in fertility and infertility: Mechanisms of action and clinical applications. *Front Endocrinol (Lausanne)*. 2022;13:1040503. doi 10.3389/fendo.2022.1040503
3. Wu L.M., Hu M.H., Tong X.H. et al. Chronic unpredictable stress decreases expression of brain-derived neurotrophic factor (BDNF) in mouse ovaries: relationship to oocytes developmental potential. *PLoS One*. 2012;7(12):e52331. doi 10.1371/journal.pone.0052331
4. Wu L.M., Liu Y.S., Tong X.H. et al. Inhibition of follicular development induced by chronic unpredictable stress is associated with growth and differentiation factor 9 and gonadotropin in mice. *Biol Reprod*. 2012;86(4):121. doi 10.1095/biolreprod.111.093468
5. Hsueh A.J., Kawamura K., Cheng Y., Fauser B.C. Intraovarian control of early folliculogenesis. *Endocr Rev*. 2015;36(1):1-24. doi 10.1210/er.2014-1020
6. Costermans N.G.J., Keijer J., van Schothorst E.M. et al. In ovaries with high or low variation in follicle size, granulosa cells of antral follicles exhibit distinct size-related processes. *Mol Hum Reprod*. 2019;25(10):614-624. doi 10.1093/molehr/gaz042
7. Caligioni C.S. Assessing reproductive status/stages in mice. *Curr Protoc Neurosci*. 2009;Appendix 4:Appendix 4I. doi 10.1002/0471142301.nsa04is48
8. Shindo M., Miyado K., Kang W., Fukami M., Miyado M. Efficient Superovulation and Egg Collection from Mice. *Bio Protoc*. 2022;12(11):e4439. doi 10.21769/BioProtoc.4439
9. Mirzayans R., Murray D. Do TUNEL and Other Apoptosis Assays Detect Cell Death in Preclinical Studies? *Int J Mol Sci*. 2020;21(23):9090. doi 10.3390/ijms21239090

doi 10.18699/bgrs2024-11.3-04

Выявление мишеней транскрипционной регуляции инсуляторного белка BEAF32 и изучение его роли в поддержании piРНК-кластеров в герминальных тканях *Drosophila melanogaster*

Соколова О.*, Кобеляцкая А., Калинкин А.
Институт биологии развития им. Н.К. Кольцова РАН, Москва, Россия
** sokolova_ol@idbras.ru*

Ключевые слова: BEAF32; инсулятор; хроматин; герминальная ткань; *Drosophila*; piРНК кластер

Мотивация и цель: Важную роль в организации структуры хроматина и регуляции экспрессии генов играют инсуляторные комплексы, или пограничные элементы генома. Инсуляторы имеют множество сайтов связывания в геноме и участвуют в сближении энхансеров и промоторов, а также в разделении доменов с разными типами хроматина. Кроме того, инсуляторные белки могут непосредственно участвовать в регуляции транскрипции, входя в состав репрессивного или активаторного транскрипционного комплекса. Инсуляторные белки млекопитающих играют важную роль в процессах гаметогенеза, а их нарушения приводят к стерильности и остановке развития. Однако, учитывая многофункциональность инсуляторных белков, зачастую сложно выявить механизмы их действия и тканеспецифичные мишени. Инсуляторный белок BEAF32 (Boundary Element-Associated Factor 32 kDa) – один из ключевых инсуляторных белков дрозофилы. Он был впервые выявлен как фактор, защищающий трансгены от влияния эффекта положения. Было отмечено, что трансгены, окруженные инсуляторными белками BEAF32, оставались транскрипционно активными при попадании в гетерохроматин.
Мутация в гене *beaf32* вызывает нарушение гаметогенеза и приводит к стерильности. В данной работе мы занимались выявлением генов-мишеней транскрипционной регуляции инсуляторного белка BEAF32 в герминальных тканях. Выявление герминальных мишеней представляет собой важную биологическую задачу. С одной стороны, это попытка определить причины самочной стерильности у мутантов; с другой – расширить понимание тканеспецифичной работы инсуляторных белков. Второе направление работы состояло в исследовании роли BEAF32 в поддержании некоторых гетерохроматиновых локусов, которые являются piРНК-кластерами, продуцирующими короткие piРНК.
Результаты: В данной работе был проведен полногеномный анализ экспрессии генов в мутантной линии BEAF32_KO в сравнении с контролем. Были использованы данные, полученные при полногеномном секвенировании библиотек кДНК из хроматиновых фракций РНК, выделенных из яичников контрольной (BEAF32/CyO) и мутантной (BEAF32_KO) линии. Графическое изображение дифференциальной экспрессии генов представлено в виде Volcano диаграммы (рис. 1).

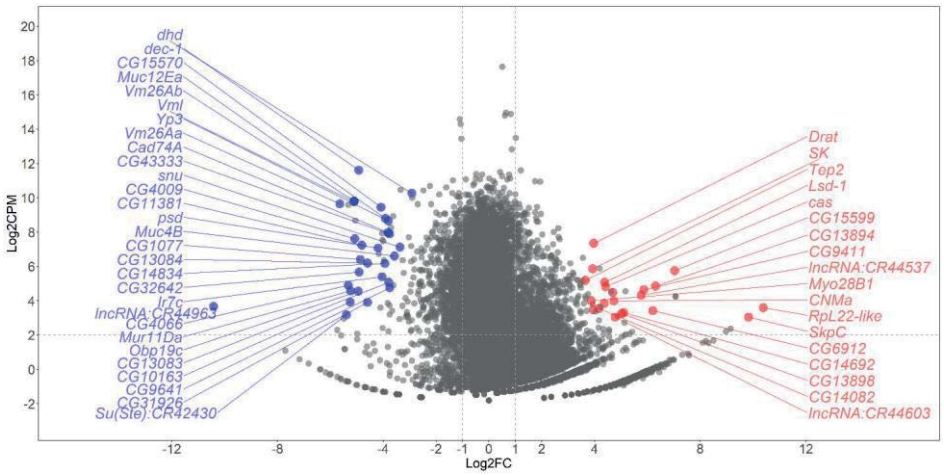


Рис. 1. Анализ дифференциальной экспрессии генов в мутантной линии BEAF32_KO в сравнении с контролем. Log_2CPM – логарифм количества прочтений, нормированный на размер библиотеки, Log_2FC – логарифм отношения значений CPM опыта к CPM контроля

Показателем, позволяющим оценить дифференциальную экспрессию, является величина Fold Change (FC). FC – это отношение количества прочтений, выровненных на аннотированную область гена, в экспериментальном образце к количеству прочтений в контрольном образце. Далее мы использовали величину log_2FC для удобства отображения данных на Volcano диаграмме. Таким образом, мы выделили две группы генов: первая группа с величиной $log_2FC > 0$, вторая группа с $log_2FC < 0$ по отношению к контролю. Для группы генов с величиной $log_2FC > 0$ было характерно увеличение экспрессии на фоне мутации *beaf32*. На Volcano диаграмме эта группа обозначена красным цветом. Для генов с величиной $log_2FC < 0$ было характерно снижение экспрессии, и на диаграмме они обозначены синим цветом. По результатам анализа был сформирован список генов с наибольшими по модулю значениями log_2FC.

Для подтверждения дифференциальной экспрессии с помощью метода обратной транскрипции с последующей ПЦР в реальном времени (RT-qPCR) были выбраны гены со значительной активацией экспрессии в яичниках на фоне BEAF32_KO и различными клеточными функциями. Дополнительным критерием для отбора генов-кандидатов являлась ассоциация их экспрессии с герминальной и нервной тканью. Для анализа мы выбрали следующие гены: *RpL22-like*, *SkpC*, *cas*, *Drat*, *Myo28B1*, *nolo*, *sug*, *CG5758*. Для подтверждения дифференциальной экспрессии методом RT-qPCR были приготовлены кДНК библиотеки из женской герминальной ткани (яичники), мужской герминальной ткани (семенники) и нервной ткани (головы). Все библиотеки кДНК были приготовлены из всех тканей мутантной и контрольной линии (рис. 2).

Все исследуемые гены показали многократный рост экспрессии в женской герминальной ткани (яичники) на фоне мутации *beaf32*, в сравнении с контролем. Значительной активации экспрессии в мужской герминальной ткани (семенники) не наблюдалось. В нейрональной ткани на фоне мутации большинство генов,

напротив, показали тенденцию к снижению экспрессии. Исключение составил ген *SkpC*, чьи значения экспрессии возрастали в сравнении с контролем. Таким образом, эффект мутации *beaf32* на экспрессию исследованных генов носит тканеспецифичный характер.

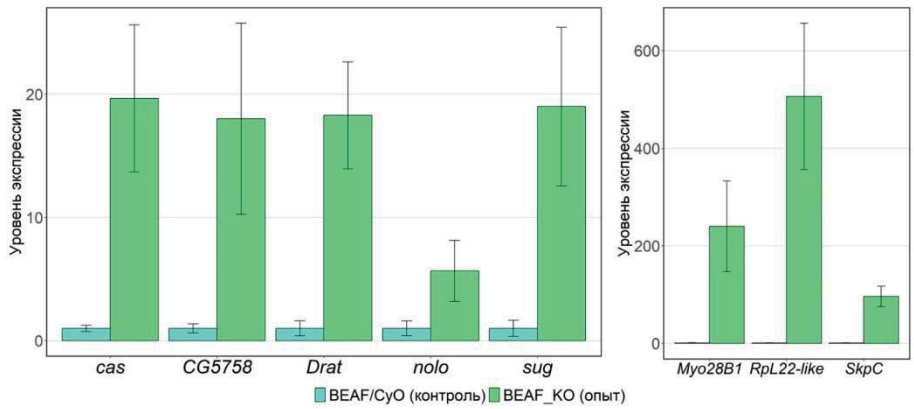


Рис. 2. Активация экспрессии генов в женской герминальной ткани (яичники) на фоне мутации *beaf32* по сравнению с контролем, определенная методом RT-qPCR. В качестве контроля загрузки использовали ген домашнего хозяйства *rp49*.
Погрешность – стандартная ошибка среднего

Далее провели анализ данных ChIP-seq библиотек и показали, что в контрольной линии в промоторной области генов *cas*, *drat*, *nolo* и *sug* собирается белковый комплекс, состоящий из BEAF32, Chriz и DREF. На фоне мутации *beaf32* белок BEAF32 не детектируется, а Chriz и DREF остаются на промоторе. Таким образом, мы делаем вывод, что для активации транскрипции вышеуказанных генов необходим инсуляторный белок BEAF32, который, с большой вероятностью, выступает как транскрипционный активатор. В промоторной области генов *RpL22-like*, *SkpC*, *Myo28B1*, экспрессия которых активируется больше чем на порядок на фоне мутации, белковый комплекс BEAF32/Chriz/DREF не детектируется. Мы предполагаем, что для данной группы генов участие BEAF32 в регуляции транскрипции происходит опосредованно, например, за счет активации экспрессии общего транскрипционного фактора.
Далее мы сосредоточились на изучении роли BEAF32 в поддержании гетерохроматиновых локусов, которые являются piРНК-кластерами. Локус *flamenco* – это одноцепочечный piРНК-кластер, который продуцирует короткие piРНК в герминальной ткани самок. Большинство протяженных piРНК-кластеров имеет гетерохроматиновую природу. Анализ данных ChIP-seq с антителами к BEAF32, Chriz и DREF показал, что в регуляторной области локуса *flamenco* на границе с активно экспрессирующимся геном DIP1 выявляется инсуляторный белок BEAF32 и его партнеры Chriz и DREF.
С помощью ChIP-qPCR мы показали, что на фоне мутации *beaf32* инсуляторный белок BEAF32 не детектируется, а его партнеры Chriz и DREF остаются в

регуляторной области *flam*. Более того, на фоне мутации транскрипция локуса *flamenco* увеличивается. Сделать однозначный вывод, за счет какой функции инсуляторного белка BEAF32 происходит активация транскрипции *flamenco*, не представляется возможным. С одной стороны, нарушение границы между доменами с разными типами хроматина может приводить к активации транскрипции, при таком сценарии реализуется барьерная функция BEAF32. С другой стороны, BEAF32 может выполнять функцию транскрипционного активатора, как и в случае некоторых генов, о которых говорилось выше.

Кроме того, анализ данных ChIP-seq показал, что инсуляторный белок BEAF32 и его партнеры Chriz и DREF локализованы между субтеломерными участками хромосом (telomere associated sequences, TAS) и эухроматиновым доменом, предположительно, формируя границу между разными типами хроматина. TAS примыкают непосредственно к теломерным повторам и, предположительно, необходимы для поддержания теломерного гетерохроматинового компартмента. Более того, субтеломерные повторы представляют собой piРНК-кластеры, которые продуцируют короткие РНК из транскриптов-предшественников. Мы показали, что на фоне мутации BEAF32 происходит уменьшение количества транскриптов-предшественников TAS. Далее мы провели эксперимент по анализу изменений количества коротких piРНК в яичниках ноль-мутантов BEAF32 в сравнении с диким типом. При мутации BEAF32 отмечалось резкое снижение количества коротких piРНК, комплементарных субтеломерным повторам (TAS). Мы не наблюдали эктопическое распространение продукции piРНК за пределы TAS, что может говорить об участии BEAF32 в транскрипционной регуляции транскриптов-предшественников, а не о формировании границ между разными типами хроматина. Изучение роли BEAF32 в герминальных тканях впервые демонстрирует важность инсуляторных белков в регуляции транскрипции тканеспецифичных генов и функционировании piРНК-кластеров, а также продукции коротких piРНК.

Финансирование: Исследование поддержано РНФ (№ 23-24-00025).

doi 10.18699/bgrs2024-11.3-05

Full-length transcriptome profiling of the rat hippocampal neurons in primary culture at different time points after activation using Nanopore sequencing

Beletskiy A.[1], Zolotar A.[1], Fortygina P.[1], Chesnokova E.[1], Uroshlev L.[1], Kolosov P.[1, 2]*

[1] *Institute of Higher Nervous Activity and Neurophysiology, RAS, Moscow, Russia*

[2] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

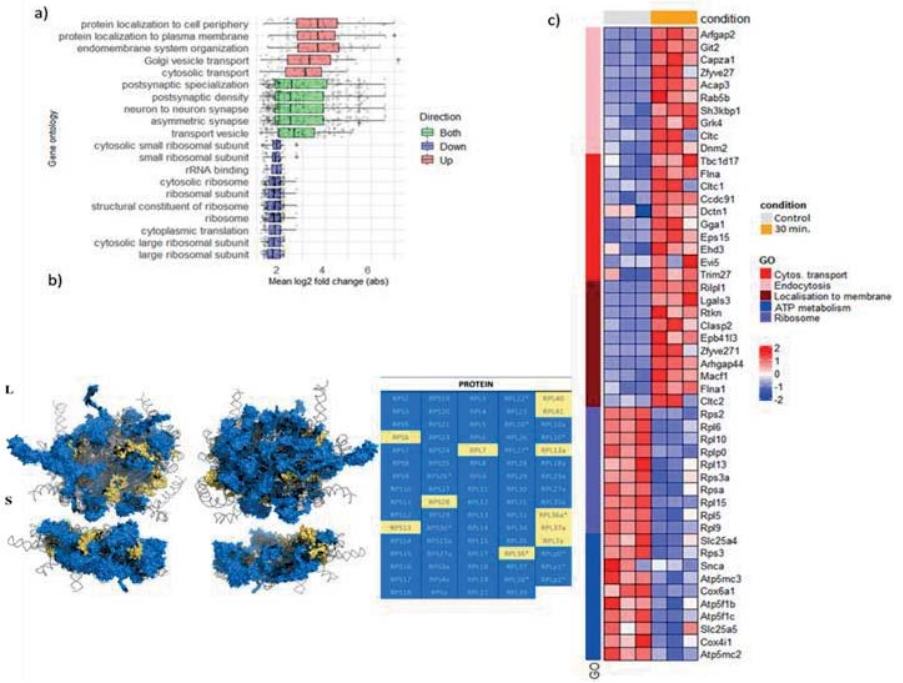
* *kolosov@ihna.ru*

Key words: long-read sequencing; ONT MinION; RNA-Seq; primary neuronal cultures; neuronal plasticity; picrotoxin; ribosomal proteins; transposable elements

Motivation and Aim: Neurons underpin cognitive ensemble processes as they are responsible for transmitting, processing, and storing information. The functional attributes are mirrored in the cell's morphology: neuronal cells are highly compartmentalized, with axons, dendrites, and the soma potentially containing distinct sets of RNA. Consequently, the transcriptional intensity and translational status of an mRNA pool can vary based on neuronal activation, extracellular conditions, and numerous other factors [1]. Neuronal stimulation can initiate the ultimate step of mRNA maturation, ensuring that mRNA synthesis is finalized just prior to the commencement of protein synthesis. The processes of transcription and translation can be separated, with various strategies in place for preserving mRNA [2]. Additionally, mRNA serves as a fundamental precursor to the protein machinery and occupies a crucial role in the execution of neuronal functions.The central objective of this study was to determine if the read count acquired through long-read sequencing with the Oxford Nanopore Technologies (ONT) MinION device is adequate to detail the transcriptomic alterations in primary neuronal cultures (PNC) following their stimulation by PTX for durations of 30 minutes, 1 hour, and 5 hours. This process is akin to the usual occurrences in a functioning brain, whereas alternative methods of neuronal activation *in vitro* tend to be less nuanced.

Methods and Algorithms: The hippocampal regions from P0 Wistar pups were prepared into primary neuronal cultures following a conventional protocol [3]. Neurons were cultured at a concentration of 250,000 cells per well. On the 15th day *in vitro* (DIV15), a standard phenol-chloroform method was used to extract RNA. Subsequently, 150 ng of total RNA was reverse transcribed with random decamer primers. In total, 20 cDNA libraries derived from poly(A)+ RNA underwent reverse transcription and were amplified into full-length cDNA using the TeloPrime Full-Length cDNA Amplification Kit V2, following the instructions provided by the manufacturer. Long-read sequencing was conducted using the MinION platform, and data production was facilitated by the ONT MinKNOW software (version 3.4.12). For the alignment of reads to the reference genome (mRatBN7), the minimap2 tool was employed. Reads aligned to the genome were quantified with FeatureCounts. The statistics for primary versus secondary alignments, as well as the proportion of reads at different coverage levels, were calculated from BAM files using the pysam library and the GenomicFeatures package in R, along with bespoke scripts. Annotation of protein domains was carried out using

the Conserved Domain Database (CDD), with a focus on PFAM domains exclusively. The coding potential was assessed using the CPAT tool. The analysis of differential expression at the gene level was conducted utilizing the DESeq2 package in R, based on transcript counts and abundances derived from Salmon. For the 30-minute mark, adjusted p-values were set at less than 0.05, and for the 1-hour and 5-hour marks, the threshold was less than 0.1. Searches within Gene Ontology and KEGG categories were carried out using the clusterProfiler package in R, applying an adjusted p-value criterion of less than 0.05. The ribosomal and endocytic pathways from the KEGG database were depicted using the Pathview online tool. The evaluation of differential transcript usage (DTU) was performed using the IsoformSwitchAnalyzer software. Four distinct approaches for assembling long-read transcriptomes were analyzed: Flair (in both relaxed and stringent modes), FLAMES, Stringtie2, and Bambu, all utilizing their respective default settings. The performance of these techniques in reconstructing transcripts at the locus level was evaluated using the software gffcompare version 0.11.7.
Results: In our study, we identified a total of 23,652 new transcripts when compared to the reference annotations, approximately 6,000 of which were completely new, predominantly originating from transposon-related locations. The examination of differentially expressed genes (DEGs) revealed that 3,046 genes exhibited differential expression; within this group, 2,037 genes showed increased expression and 1,009 showed decreased expression 30 minutes following PTX treatment. At subsequent time points of 1 hour and 5 hours, the numbers of differentially expressed genes were significantly lower, with 446 and 13 genes respectively (see the Fig. 1a–c).
Conclusion: Gene Ontology and KEGG pathway analyses indicated a direct correlation between both upregulated and downregulated genes with protein metabolism processes. Significantly, genes involved in protein transport and localization saw an increase in regulation, whereas numerous genes responsible for ribosomal proteins, which normally have a high expression level, were downregulated following a 30-minute incubation with PTX. This suggests a reallocation of transcriptional resources in favor of genes activated by the treatment. For the ribosome-associated proteins Rpl15 and Rps8, a marked decrease in expression was observed in neuronal cultures after a 30-minute exposure to PTX, as verified by digital PCR (dPCR; Fig.1d).
We observed a significant reduction in the expression of genes involved in ATP metabolism, predominantly those associated with mitochondrial genes that are typically highly expressed. Additionally, there was a substantial alteration in the expression levels of genes within the endocytosis pathway, with the majority of these genes experiencing upregulation. A 30-minute incubation with PTX did not result in a significant increase in alternative splicing events. Surprisingly, only slight variations were observed at each time point of the experiment (with 16, 16, and 4 isoforms affected after 30 minutes, 60 minutes, and 5 hours, respectively), according to Differential Transcript Usage (DTU) analysis. Nevertheless, there was a prominent switch in the transcripts of the Gabrb1 gene, which is responsible for encoding the β1 subunit of the GABA receptor, a known direct target of PTX.



d)

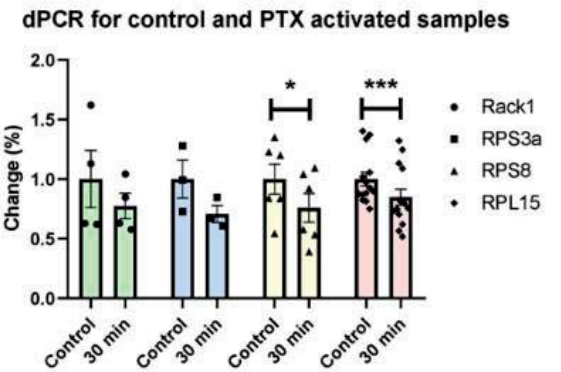


Fig. 1. a) Top 10 down- and up-regulated Gene Ontology categories for DEGs identified at 30 min after the PTX application. Y axis represents GO category, X axis – log2 fold change, averaged across all genes in the corresponding category. b) Visualization of the ribosomal pathway genes that are differentially expressed at 30 min after the PTX application. Structure of the mammalian ribosome, two views each for the large (L) and small (S) subunits. Grey color represents rRNA. Protein products of downregulated genes are shown in blue, protein products of non-DE genes are shown in yellow. c) Heatmaps showing top 10 DEG (sorted by log 2 fold change) per GO category. d) dPCR results for selected genes encoding ribosome and ribosome-associated proteins. Control group: $n = 4$ for Rack1; $n = 3$ for Rps3a; $n = 6$ for Rps8; $n = 15$ for Rpl15; "30 min" PTX activation group: $n = 4$ for Rack1; $n = 3$ for Rps3a; $n = 6$ for Rps8; $n = 15$ for Rpl15 biological replicates. Expression levels of genes of interest were normalized to the average Hprt expression in the same sample. The results are present as proportional change in gene expression ± SEM normalized to the control group value; * – $p < 0.05$, *** – $p < 0.001$, Wilcoxon signed rank test

The discovery of new loci and isoforms in this research could enhance our comprehension of the functional mRNA array involved in neuronal plasticity. Concurrently, the gene clusters annotated by Gene Ontology (GO) could provide deeper insights into the specific molecular mechanisms that explain the absence of inhibitory signals from interneurons, which is a result of the blockade of GABA receptors during normal physiological functions.

Funding: This work was funded by the Russian Science Foundation, agreement No. 23-14-00331.

References

1. Cho J. et al. Multiple repressive mechanisms in the hippocampus during memory formation. *Science*. 2015;350(6256):82-87. doi 10.1126/science.aac7368
2. Chesnokova E.A., Kolosov P.M. Local Protein Synthesis in Dendritic Terminals and Its Regulation in Normal Conditions and during Plastic Changes. *Neurosci Behav Physi.* 2017;47:595-607. doi 10.1007/s11055-017-0440-0
3. Sessegolo C. et al. Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA molecules. *Sci Rep*. 2019;9(1):14908. doi 10.1038/s41598-019-51470-9

doi 10.18699/bgrs2024-11.3-06

Linkage of alternatively spliced exons proliferates in neuronal genes

Boltunov T.[1], Babenko V.[2]*

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

* *bob@bionet.nsc.ru*

Key words: AS exons co-variation; coordinated alternative splicing within a gene; AS exons cluster; AS mediated information complexity

Based on publicly available RNA-seq data of human hippocampus samples, we identified alternatively spliced (AS) exons genome wide along with assessing the genes percentage spliced in (psi) values. The data has been compiled on more than 30 samples for each gene. Along with psi values we compiled pairwise covariation matrix across all AS (exon skipping) exons based on Pearson r2. Further Agglomerative Hierarchical Clustering (AHC) procedure on matrix revealed dense AS exons clusters with linked exons, with clusters size in the interval of [2…19] (pairwise correlation p-value < 1E-6). There were around 2,200 genes with clusters of coordinated AS exons.

Further analysis revealed AS clusters maintain antagonistic or independent relations within a gene. We explored the traits of genes abundant with AS clusters: the majority proved to be neurospecific genes, including synapse, and cytoskeletal (axonal) genes. Notably, Neurospecific splice factors (SFs) also maintain expanded coordinated AS regulation.

While the previous study observed coordinated splicing before [1], the scale of the phenomenon has not been explicitly highlighted. From the evolutionary point of view, and, given the information complexity of splicing decreases upon exon co-variation, we may speculate that the rapid response to the homeostatic environment favors quick coordinated splicing mediated tune-up of the gene's isoform. Still, the mechanistic background of phenomenon is not elucidated. One of the viable hypothesis is the specific secondary structure of mRNA favoring the quick coordinated SFs binding.

In our report we address the phenomenon and consider several examples of coordinated AS.

Funding: This work has been supported by ICG SB RAS state program FWNR-2022-0020.

References

1. Fagnani M. et al. Functional coordination of alternative splicing in the mammalian central nervous system. *Genome Biol*. 2007;8(6):R108

doi 10.18699/bgrs2024-11.3-07

Molecular mechanisms of cell death under cold plasma jet exposure: potential enhancement with gold nanoparticles

Koval O.[1, 2]*, Biryukov M.[1, 2, 3], Polyakova A.[1, 2, 3], Krychkova N.[1, 2], Semenov D.[1], Gorbunova E.[1, 2], Pyshnaya I.[1], Zakrevsky Dm.[2, 4], Milakhina E.[2, 4], Schweigert I.[2]

[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*
[2] *Khristianovich Institute of Theoretical and Applied Mechanics, SB RAS, Novosibirsk, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
[4] *A.V. Rzhanov Institute of Semiconductor Physics, SB RAS, Novosibirsk, Russia*
* *o.koval@niboch.nsc.ru*

Key words: cold plasma jet; cytotoxic activity; apoptosis; cell cycle

Motivation and Aim: Cold physical plasma (CP) is a new modality in cancer research that can help enhance the results of chemotherapy and immunotherapy. The result of plasma interaction with biological object is the generation of ions, oxygen- and nitrogen-related species (RONS) in the gas and water environment that actively react with the cells of biological tissue, causing cell death under certain irradiation regimes. The sensitivity of various healthy and tumor cells to plasma irradiation has been shown to be different. Therefore, the study of the mechanisms of resistance of grown human cells to the cold plasma jet of atmospheric pressure and the search for overcoming such resistance is an urgent task of plasma medicine.

Methods and Algorithms: The helium plasma jet was initiated with a generator of sinusoidal voltage with the amplitude of (3.3–3.5) kV at a fixed frequency f_U = 50 kHz and with a generator of unipolar positive pulses with adjustable repetition rate f = 1–40 kHz and amplitude of 3.8 kV and 4.2 kV. Under these modes of plasma jet generation, the near-selective conditions of tumor cell death were determined in *in vitro* experiments. In irradiated cells, dynamic changes in the transcriptome were assessed using the RNA-Seq (Illumina 1500 NextSeq platform) and subsequent gene sets enrichment analysis. Data verification for the particular transcripts was done by RT-PCR and western blot for corresponding proteins. Gold nanoparticles (GNP) were synthesized and covered with polyethylene glycol (PEG) for the co-treatment cells with cold plasma exposure.

Results: The adding of GNP to the CAP-treated cells increased cell death rate compared to cells treated with plasma alone. Based on the transcriptome-analysis data, the genes KLF4, FOS, ATF3, and GADD45B were selected to confirm changes in their mRNA levels in response to CAP exposure and combinatorial treatment of CAP with GNP. The results of the RT-PCR assay showed that the expressions of the *KLF4*, *FOS*, and *ATF3* mRNAs were up-regulated 3 h after the treatment, and GNP did not alter the level and type of transcriptional response. Such a transcriptional response is implicated in the p53 pathway, KRAS signaling, UV response, TNF-alpha signaling, and apoptosis-related processes. The amplitude of the response to CAP treatment was more variable in the A549 cells.

Conclusion: The combination of gold nanoparticles with CAP exposure was a quite successful in the *in vitro* experiments against various cancer cells. We suppose that this way is a promising strategy for treatment of cancerous *in vivo*, especially using modified GNP.

Funding: The study is supported by the RSF grant No. 22-49-08003.

doi 10.18699/bgrs2024-11.3-08

Analysis of apoptotic activity of *Helichrysum arenarium* extract *in vitro*

Polukonova N., Kurchatova M.*, Navolokin N., Baryshnikova M., Mylnikov A., Polukonova A., Durnova N.

Saratov State Medical University named after V.I. Razumovsky, Saratov, Russia

** kurchatova.marya@yandex.ru*

Key words: apoptosis; flavonoids; kidney cancer

Motivation and Aim: Sandy immortelle (*Helichrysum arenarium* (L.) Moench) is a perennial herbaceous plant of the Asteraceae family, growing in Europe and Central Asia. An infusion of immortelle flowers is used as a choleretic agent. It is known that immortelle flowers contain a significant amount of flavonoids (at least 20 compounds) belonging to various groups. Today, flavonoids are one of the most promising substances of plant origin for the creation of drugs, including those for the fight against cancer. A previous *in vivo* study of immortelle sandy extract on rats with transplanted sarcoma 45 established the positive effect of the extract in the treatment of malignant neoplasms: a decrease in tumor volume, as well as necrotic and dystrophic processes in it. The purpose of the work was to study the possible apoptotic activity of a flavonoid-containing extract of immortelle sandy on a culture of human kidney cancer cells A498.
Methods and Algorithms: The material for preparing the extracts was the flowers of the sandy immortelle. Immortelle extract was obtained according to the method described in Patent for invention No. 2482863: "Method of obtaining a dry extract from plant materials with biological activity". A study of the antitumor activity of immortelle extract was carried out on human kidney cancer A498 tumor cells obtained from the bank of tumor cultures of the Russian Cancer Research Center named after. N.N. Blokhina. Cell cultivation was carried out in plastic bottles in RPMI 4. Cells were cultured in a CO_2 incubator at 37 °C for 24 hours, after which they were stained. A Nikon microscope was used to visualize the cells. ImageJ software was used for cell counting. To detect morphofunctional changes in cells, staining with two dyes was used: acridine orange, which stains living cells, and propidium iodide, which penetrates non-viable cells.
Results: We studied the following concentrations of immortelle extract (mg/ml): 0.9; 1.8; 3.6; 7.2. The tests were carried out after 24 and 48 hours. A comparison was made in the control and experimental groups according to the following indicators: the number and ratio of cells with sickles to the number of living cells, the number and ratio of cells with pyknosis to the number of living cells, the number and ratio of cells in apoptosis to the number of living cells cells. The data were then statistically analyzed using SPSS 17.0 software. The normality of the distribution of characteristics in groups was determined using the Shapiro–Wilk test. The distribution of characteristics did not correspond to normal, therefore, to assess the presence of statistically significant differences, the nonparametric Mann–Whitney test (U-test) was used. Quantitative data were described using median, minimum and maximum values, 1st and 3rd quartiles. A result was considered statistically significant if the probability of rejecting the null hypothesis of no differences did not exceed 5 % ($p < 0.05$). Such indicators of apoptotic

activity as the number of cells with sickle-shaped nuclei, cells with pyknosis of the nucleus, as well as the ratio of the number of such cells to the number of living cells did not reveal differences between cells under the influence of immortelle extract and control. However, the number of cells that disintegrated into apoptotic bodies under the influence of immortelle extract at concentrations from 0.9 to 3.6 mg/mg after 24 hours was slightly higher than in the control, which indicates the presence of weak apoptotic activity of the extract as a whole. Weak apoptotic activity was also detected after 48 hours on the second day: at an extract concentration of 0.9 mg/ml, after 48 hours there is an increase in cells that disintegrated into apoptotic bodies, compared with 24 hours.
Conclusion: Thus, the extract of immortelle sandy has a weak apoptotic effect on the culture of human kidney cancer cells A498.
Funding: The study is supported within the framework of the R&D initiative project "Obtaining and studying plant extracts and establishing the influence of their chemical composition on biological activity" (AAA-A20-120011590047-2).

doi 10.18699/bgrs2024-11.3-09

Functional cross-talks between transcription, alternative splicing, translation and novel post-translational modifications in Alzheimer disease: fill missings using proteomics

Tarasova I.[1]*, Brazhnikov M.[1, 2], Kopeykina A.[1, 3], Emekeeva D.[1, 4]

[1] *V.L. Talrose Institute for Energy Problems of Chemical Physics, N.N. Semenov Federal Research Center of Chemical Physics, RAS, Moscow, Russia*

[2] *Osipyan Institute of Solid State Physics, RAS, Chernogolovka, Russia*

[3] *Pirogov Russian National Research Medical University, Moscow, Russia*

[4] *Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia*

* *iatarasova@yandex.ru*

Key words: Alzheimer disease; post translational modifications; amino acid substitutions; alternative splicing; molecular mechanisms

Motivation and Aim: Alzheimer's disease (AD) is the most common dementia in clinical practice. According to WHO, the number of AD cases worldwide was 55 million in 2022, and the forecasts promise an increase by 10 million every year (https://www.who.int/news-room/fact-sheets/detail/dementia). Despite the research community contributing significant efforts in studying pathogenesis, understanding the molecular and structural changes in the cells remains incomplete. For example, alternative splicing (AS) represents the powerful mechanism of regulating protein function. However, the diversity of alternatively spliced proteins and their role in pathogenesis of Alzheimer disease are not investigated. Detection of AS at protein level is a challenging task, consequently, the contribution of alternatively spliced proteins to the cell life cycle is basically unknown and represents a matter of contradictory statements from no impact on proteome diversity to evidences that most frame-preserving alternatively spliced isoforms are translated. There are three major factors complicating measurement of alternative splicing by standard methods of mass spectrometry based proteomics: (1) low protein sequence coverage (30 % in average); (2) low proteome coverage (products from the best case of 1/2 to the regular 1/6 of human protein coding genes); (3) approximately 1/2 of known isoforms can be identified by the only one tryptic peptide from the alternatively spliced part of protein sequence. Next level of complexity is identification and characterization of post translational modifications (PTM). Most approaches assume characterization of a targeted PTM enriched by a specific experimental protocol. However, post translational modifications trigger other PTMs and together they provide the regulation of protein function. Investigation of such events requires simultaneous measurements of different types of PTMs. We argue that big proteomics data analysis is of high potential to make a first step to PTM interplay discoveries as well as for discovery of new PTMs affecting protein functions in pathology development. Our motivation behind the study is to develop mass spectrometry based bioinformatic approaches for discovery of the alternative splicing and amino acid substitutions translated into proteins in Alzheimer's disease as well as novel post translational modifications including PTMs of the alternatively spliced regions and characterization of their functional implications.

Methods and Algorithms: Public proteomics data was used in the study [1]. Proteomics data collected by the Mount Sinai Brain Bank (MSBB) (https://doi.org/10.7303/syn5759470) and the Banner Sun Health Research Institute (Banner) (https://doi.org/10.7303/syn7170616) were downloaded from AD Knowledge Portal at http://www.synapse.org within the controlled-access data use certificate. In total, the *post mortem* brain tissues (prefrontal cortex) from 300 AD patients and 200 healthy individuals were analyzed. Protein identification was performed using different strategies and their combinations: database search (IdentiPy [2], MSFragger [3]) within the strict and loose mass tolerances [4], and database search with MSFragger search engine supplemented by *de novo* peptide sequence assembling using Novor software [5]. Database search was performed against the canonical and isoform UniProt TrEMBL protein database (taxon ID: 9604) concatenated with a decoy database created by reversing the protein sequences from the UniProt TrEMBL database. Post search validation was performed using Scavager software [6]. Positions of post translational modifications were established using AA_stat [4]. Differential expression of splice protein isoforms was assessed based on the normalized intensities of peptide ions in the mass spectra reported by TMTcrunch software developed in this study and calculation of standard mean deviations (SMD). Statistical analysis was performed using SciPy [7] and program code developed in house. Gene ontology enrichment analysis was performed using STRING [8]. To validate variant peptides, the chromatographic retention time (RT) prediction with deep machine learning was used. Peptides with ΔRT value within $\pm\sigma$ of the median value were considered true. Pathogenicity of the identified amino acid polymorphisms was evaluated using VarSome [9]. UniProt and InterPro databases were used to match alternative splicing events and post translational modifications falling into the protein functional domains and binding sites.
Results: Three pipelines for identification of alternatively spliced proteins, amino acid substitutions and post translational modifications were developed. First pipeline allowed us to report translation of approximately 1,300 splice isoforms presented in the SwissProt database, unambiguously identified by isoform-unique peptide features. We observed translation of alternative splicing affecting functional domains and binding sites for ASPH, ATP2B2, DLGAP1, DNM1L, MAPT and other AD-related protein coding genes. Second pipeline implements identification of tryptic peptides by the *de novo* sequencing from mass spectrometry data, database search against the SwissProt canonical database, and comparison of peptides assembled *de novo* and peptides identified in database search to find out the peptide sequences differing by amino acid substitution, insertion or deletion. The pipeline provides a competitive search between variant and canonical peptides and produces tables of reliable peptides with amino acid substitutions, their positions in the protein, the corresponding canonical peptides and protein names. Using the developed workflows we identified a representative group of amino acid substitutions annotated in the Unimod database as misacylation of tRNA. Such events result in tRNA-dependent mistranslation that is earlier reported in microorganisms and human cells [10]. In our results, we observed tRNA mistranslation for the dihydropyrimidinase-like (DPYSL) proteins, the Na,K-ATPases ATP1A1/2/3, calcium/calmodulin-dependent kinase 2 and other proteins involved in neuro-developmental disorders. Our next findings were amino acid substitutions corresponding to single nucleotide polymorphisms that were predicted and/or annotated as pathogenic. In particular, we found proteomics evidences for a number of known pathogenic substitutions: HBB Q128E, GFAP N77D, NEFL N98D, etc. Using the developed

workflows we characterized the landscape of post translational modifications in precortex tissues of AD patients and healthy individuals. It was represented by deamidation of asparagine, glutamine and arginine, dehydrogenation, acetylation, phosphorylation, oxidation, glycosylation, methylation and many others. In total we identified few dozen of proteins with statistically significant changes in the level of post translationally modified proteins in Alzheimer's disease relative to the control group. Cross-validation using different bioinformatic approaches and comparison with deamidation and citrullination annotated in UniProt and dbPTM allowed us to propose three dozen of new deamidation sites enriched in AD and located on N, Q and R in protein products of MYO5A, ATP1A3, GATD3B, NEFM, and TKT genes. In analysis of amino acid motifs, the *N*G motif was enriched in peptides with modified asparagine.
Conclusion: The developed bioinformatic pipelines for processing mass spectrometry data represent powerful tools to follow genomic events translated to proteins. Using those workflows, we observed numerous molecular events pretending for reproducibility and novelty, and their analysis is to be continued. Our efforts are further focused on data integration, interpretation and systematization to reconstruct the molecular mechanisms.
Funding: The study is supported by the Russian Science Foundation (No. 23-45-00012).

References

1. Bereczki E., Branca R.M., Francis P.T. et al. Synaptic markers of cognitive decline in neurodegenerative diseases: a proteomic approach. *Brain.* 2018;141(2):582-595. doi 10.1093/brain/awx352
2. Levitsky L.I., Ivanov M.V., Lobas A.A. et al. IdentiPy: an extensible search engine for protein identification in shotgun proteomics. *J Proteome Res.* 2018;17(7):2249-2255. doi 10.1021/acs.jproteome.7b00640
3. Kong A.T., Leprevost F.V., Avtonomov D.M., Mellacheruvu D., Nesvizhskii A.I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry – based proteomics. *Nat Methods.* 2017;14(5):513-520. doi 10.1038/nmeth.4256
4. Levitsky L.I., Bubis J.A., Gorshkov M.V., Tarasova I.A. AA_stat: Intelligent profiling of *in vivo* and *in vitro* modifications from open search results. *J Proteomics.* 2021;248:104350. doi 10.1016/j.jprot.2021.104350
5. Ma B. Novor: Real-time peptide *de novo* sequencing software. *J Am Soc Mass Spectrom.* 2015;26(11):1885-1894. doi 10.1007/s13361-015-1204-0
6. Ivanov M.V., Levitsky L.I., Bubis J.A., Gorshkov M.V. Scavager: a versatile postsearch validation algorithm for shotgun proteomics based on gradient boosting. *Proteomics.* 2019;19(3):e1800280. doi 10.1002/pmic.201800280
7. Virtanen P., Gommers R., Oliphant T.E. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* 2020;17(3):261-272. doi 10.1038/s41592-019-0686-2
8. Szklarczyk D., Gable A.L., Nastou K.C. et al. The STRING database in 2021: customizable protein – protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;49(D1):D605-D612. doi 10.1093/nar/gkaa1074
9. Kopanos C., Tsiolkas V., Kouris A., Chapple C.E., Albarca Aguilera M., Meyer R., Massouras A. VarSome: the human genomic variant search engine. *Bioinformatics.* 2019;35(11):1978-1980. doi 10.1093/bioinformatics/bty897
10. Lant J.T., Berg M.D., Sze D.H.W. et al. Visualizing tRNA-dependent mistranslation in human cells. *RNA Biol.* 2018;15(4-5):567-575. doi 10.1080/15476286.2017.1379645

12 Симпозиум «Математические проблемы биоинформатики и системной компьютерной биологии. Анализ больших генетических данных и искусственный интеллект»

Symposium “Mathematical problems of bioinformatics and systems computational biology. Big genetic data analysis and artificial intelligence”

doi 10.18699/bgrs2024-12.1-01

Компьютерное моделирование биохимических процессов фиброза печени при помощи машинного обучения, информированного физикой

Бондарева А.К.[1]*, Кириллов Б.А.[2, 3]

[1] *Московский физико-технологический институт, Долгопрудный, Россия*

[2] *Центр технологий материалов, Сколковский институт науки и технологий, Москва, Россия*

[3] *Центр высокоточного редактирования и генетических технологий для биомедицины, Институт биологии гена РАН, Москва, Россия*

* *bondareva.ak@phystech.edu*

Ключевые слова: физико-информированная модель; фиброз печени; математическое моделирование; прямые биомаркеры; машинное обучение; дифференциальные уравнения

Мотивация и цель: Патологии печени оказывают значительное влияние на здравоохранение из-за их высокой распространенности и отрицательного воздействия на экономику [1]. Диагностика и лечение заболеваний печени усложняются их асимптоматическим течением и ограниченностью доступных данных о прямых биомаркерах, включая те, что получены инвазивными методами [2]. Применение физико-информированного машинного обучения позволяет преодолеть существующие ограничения, помогая оптимизировать исследовательский процесс, сократить вычислительные затраты и улучшить предсказательную способность моделей, таким образом открывая новые возможности для глубокого понимания механизмов заболевания и оценки эффективности терапевтических вмешательств.

Методы и алгоритмы: Для решения задачи моделирования динамики фиброза печени используется подход физико-информированного машинного обучения, сочетающий алгоритмы машинного обучения с моделями, основанными на методах математической физики. Основой подхода служит интеграция системы из 24 дифференциальных уравнений в алгоритм машинного обучения для точного описания биохимических процессов, происходящих в биологической системе [3]. Строгое соблюдение физических законов обеспечивается встраиванием дополнительных ограничений в функцию потерь, минимизируемую в процессе обучения модели [4].

Результаты: Предлагается *in silico* модель динамики процессов фиброза печени, основанная на решении методами машинного обучения системы дифференциальных уравнений, описывающей физические и биохимические процессы фиброгенеза. Применение физико-информированного машинного обучения позволяет снизить требования к вычислительным ресурсам, делая процесс моделирования более доступным и экономичным.

Выводы: Разработанная модель фиброгенного процесса в печени на основе физико-информированного машинного обучения позволяет воспроизвести динамику заболевания для помощи в принятии медицинских решений. Результаты исследования позволяют глубже понимать механизмы развития фиброза для оптимизации разработки диагностических и лечебных протоколов.

In silico modeling for biochemical processes of liver fibrosis based on physics-informed machine learning

Bondareva A.[1]*, Kirillov B.[2, 3]

[1] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

[2] *Center of Material Technologies, Skolkovo Institute of Science and Technology, Moscow, Russia*

[3] *Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, RAS, Moscow, Russia*

** bondareva.ak@phystech.edu*

Key words: PIML; liver fibrosis; mathematical modeling; direct biomarkers; machine learning; differential equations

Motivation and Aim: Liver pathologies significantly impact healthcare due to their high prevalence and negative effects on economics [1]. Unfortunately, asymptomatic progression and limited availability of data about direct biomarkers (including the data obtained through invasive methods [2]) complicate diagnosis and treatment of liver diseases. Application of physics-informed machine learning offers a promising approach to overcome these limitations by optimizing the research process, reducing computational costs, and enhancing the predictive accuracy of models, opening up new possibilities for a deeper understanding of disease mechanisms and assessing the efficacy of therapeutic interventions.

Methods and Algorithms: We use a physics-informed machine learning approach to model liver fibrosis dynamics. Our approach is based on integration of machine learning algorithms with methods of mathematical physics – incorporation of a system of 24 differential equations into the machine learning model, thus enabling precise description of biochemical processes within the biological system of interest [3]. Strict adherence to physical laws is achieved by introduction of additional constraints into the model's loss function, minimized during the training of the model [4].

Results: We propose an *in silico* model of liver fibrosis dynamics based on application of machine learning methods for solving a system of differential equations that describes physical and biochemical processes of fibrogenesis. The use of physics-informed machine learning reduces the requirements for computational resources, thus making the modeling process more accessible and cost-effective.

Conclusion: The physics-informed machine learning model of fibrogenic process in the liver allows for reproduction of disease dynamics for the needs of medical decision making assistance. Results of the research provide a deeper understanding of fibrosis evolution for the development of diagnostic and therapeutic protocols.

Список литературы/References

1. Devarbhavi H. et al. Global burden of liver disease: 2023 update. *J Hepatol*. 2023;79(2):516-537. doi 10.1016/j.jhep.2023.03.017
2. Iredale J.P. Models of liver fibrosis: exploring the dynamic nature of inflammation and repair in a solid organ. *J Clin Invest*. 2007;117(3):539-548. doi 10.1172/JCI30542
3. Friedman A., Hao W. Mathematical modeling of liver fibrosis. *Math Biosci Eng.* 2017;14(1):143-164. doi 10.3934/mbe.2017010
4. Karniadakis G.E. et al. Physics-informed machine learning. *Nat Rev Physic*s. 2021;3(6):422-440. doi 10.1038/s42254-021-00314-5

doi 10.18699/bgrs2024-12.1-02

Гемодинамика аневризм: моделирование и эксперимент

Паршин Д.В.[1]*, Липовка А.И.[1], Куянова Ю.О.[1], Тихвинский Д.В. [1],
Карпенко А.А.[2], Дубовой А.В.[3], Бервицкий А.В.[3] Бесов А.С.[1], Чупахин А.П.[1]

[1] *Институт гидродинамики им. М.А. Лаврентьева СО РАН, Новосибирск, Россия*

[2] *НМИЦ им. ак. Е.Н. Мешалкина, Новосибирск, Россия*

[3] *Федеральный нейрохирургический центр, Новосибирск, Россия*

* *parshin@hydro.nsc.ru*

Ключевые слова: аневризма; рисковые критерии; виртуальная хирургия; численная гемодинамика; прочностные свойства; вязкоупругие свойства

Мотивация и цель: По данным ВОЗ [1] сосудистые патологии занимают лидирующее место как по количеству населения, так и по смертности по всему миру. Аневризмы составляют значительную часть таких патологий и могут протекать как с проявлениями, так и без проявления симптомов. В современной высокотехнологичной персонализированной медицине большая роль отводится рисковой характеристике аневризм: формирования прогнозов роста и разрыва аневризм. Такая характеристика невозможна без понимания фундаментальных основ их гидродинамики, а также механики тканей аневризм.

Методы и алгоритмы: В данной работе приводятся конкретные подходы, позволяющие сформировать рисковые критерии наблюдения, операции и возникновения постоперационных осложнений для церебральных аневризм (ЦА) и аневризмы брюшной аорты (АБА). Для определения фундаментальных основ гидродинамики аневризм использовался экспериментальный подход с формированием лабораторного гидродинамического контура. Для определения механических свойств использовались различные экспериментальные техники: одноосный механический тест, измерения вязкоупругих свойств, лазерно-индуцированная флуоресценция. Для предсказательного анализа использовалось численное моделирование в ANSYS, а также статистический анализ полученных результатов.

Результаты: Для церебральных аневризм удалось классифицировать материал аневризм согласно их статусу. Оказалось, что в зависимости от статуса церебральных аневризм существенно по-разному экспериментальные данные описываются гиперупругими моделями материалов, которые наиболее часто применяются в вычислительных пакетах. Дана подробная классификация применения таких моделей как в зависимости от статуса аневризмы, так и в зависимости от величины деформации ткани аневризмы [2]. Кроме того, методами ЛИФ и статистического анализ данных одноосного теста удалось показать наличие корреляционной связи между данными таких экспериментов. Оказалось, что, вновь, в зависимости от статуса аневризмы возможно построить модели линейной регрессии между энергетическим откликом на излучение и прочностными характеристиками. Подобные регрессии возможны как для величин предельного напряжения, так и для величин предельной деформации [3]. Методами численного моделирования было показано, что если оценивать разницу величины пристеночных сдвиговых напряжений в области за виртуально-

установленным стентом, то возможно классифицировать по такой величине предполагаемый операционный исход по градации: нет окклюзии, стеноз в области стента, удачная окклюзия [4].
Для АБА с помощью численного моделирования удалось показать прямую связь между ростом аневризмы и ростом давления в системе кровообращения проксимально аневризме при особой конфигурации бифуркации аорты и расположения аневризмы на бифуркации. Это означает, что наблюдение пациентов со схожей конфигурацией бифуркации и бифуркационным положением аневризмы должно происходить регулярнее во избежание возникновения сопутствующих сердечных патологий [5]. Кроме того, с помощью решения контактной задачи о механике внутрипросветного тромба в АБА удалось показать влияние асимметрии на механику стенки и выявить механизм такого влияния [6].
Выводы: Представленные результаты позволяют решать конкретные клинические задачи по наблюдению, операции и оценке постоперационных рисков пациентов с церебральными аневризмами и аневризмами брюшной аорты. Кроме того, в ходе реализации работы получены интересные подходы классификации математических моделей, разработки гемодинамических контуров и подходов к анализу результатов вязкоупругих тестов.
Финансирование: Работа выполнена при финансовой поддержке гранта Российского научного фонда (проект 20-71-10034, https://rscf.ru/project/20-71-10034/).

Hemodynamics of aneurysms: modeling and experiment

Parshin D.V.[1]*, Lipovka A.I..[1], Kuianova Iu.O.[1], Tikhvinsky D.V.[1], Karpenko A.A.[2], Dubovoi A.V.[3], Bervitsky A.V.[3], Besov A.S.[1], Chupakhin A.P.[1]

[1] *Lavrentyev Institute of Hydrodynamics, SB RAS, Novosibirsk, Russia*

[2] *Meshalkin Research Center of Circulation Pathology, Novosibirsk, Russia*

[3] *Federal Neurosurgical Center, Novosibirsk, Russia*

* *parshin@hydro.nsc.ru*

Key words: aneurysm; risk criteria; virtual surgery; numerical hemodynamics; strength properties; viscoelastic properties

Motivation and Aim: According to WHO [1], vascular pathologies occupy a leading place both in terms of population and mortality throughout the world. Aneurysms make up a significant part of such pathologies and can occur with or without symptoms. In modern high-tech personalized medicine, a large role is given to the risk characteristics of aneurysms: the formation of forecasts for the growth and rupture of aneurysms. Such characterization is impossible without understanding the fundamental principles of their hydrodynamics, as well as the mechanics of aneurysm tissue.
Methods and Algorithms: This paper provides specific approaches that allow us to formulate risk criteria for observation, surgery, and the occurrence of postoperative complications for cerebral aneurysms (CA) and abdominal aortic aneurysm (AAA). To determine the fundamental principles of the hydrodynamics of aneurysms, an experimental approach was used with the formation of a laboratory hydrodynamic contour. To determine the mechanical properties, various experimental techniques were

used: uniaxial mechanical test, viscoelastic properties measurements, laser-induced fluorescence. Numerical modeling in ANSYS was used for predictive analysis.
Results: For cerebral aneurysms, it was possible to classify aneurysm material according to their status. It turned out that, depending on the status of cerebral aneurysms, experimental data are described significantly differently by hyperelastic material models, which are most often used in computing packages. A detailed classification of the use of such models is given [2], both depending on the status of the aneurysm and depending on the amount of deformation of the aneurysm tissue. In addition, using LIF methods and statistical analysis of uniaxial test data, it was possible to show the presence of a correlation between the data of such experiments. It turned out that, again, depending on the status of the aneurysm, it was possible to construct linear regression models between the energy response to radiation and strength characteristics. Similar regressions are possible for both the ultimate stress values and the ultimate strain values [3]. Using numerical modeling methods, it was shown that if we evaluate the difference in the magnitude of near-wall shear stresses in the area behind the virtually installed stent, then it is possible to classify the expected surgical outcome according to this value according to gradation: no occlusion, stenosis in the stent area, successful occlusion [4]. For AAA, using numerical modeling, it was possible to show a direct relationship between the growth of the aneurysm and the increase in pressure in the circulatory system proximal to the aneurysm for a special configuration of the aortic bifurcation and the location of the aneurysm on the bifurcation [5]. This means that patients with a similar bifurcation configuration and bifurcation position of the aneurysm should be monitored more regularly to avoid the occurrence of concomitant cardiac pathologies. In addition, by solving the contact problem on the mechanics of an intraluminal thrombus in an AAA, it was possible to show the influence of asymmetry on the mechanics of the wall and identify the mechanism of such an influence [6].
Conclusion: The presented results make it possible to solve specific clinical problems in monitoring, surgery and assessing postoperative risks of patients with cerebral aneurysms and abdominal aortic aneurysms. In addition, during the implementation of the work, interesting approaches were obtained for classifying mathematical models, developing hemodynamic contours and approaches to analyzing the results of viscoelastic tests.
Funding: The work was carried out with financial support from a grant from the Russian Science Foundation (project 20-71-10034, https://rscf.ru/project/20-71-10034/).

Список литературы/References

1. https://www.who.int/ru
2. Parshin D.V., Lipovka A.I., Yunoshev A.S., Ovsyannikov K.S., Dubovoy A.V., Chupakhin A.P. On the optimal choice of a hyperelastic model of ruptured and unruptured cerebral aneurysm. *Sci Rep.* 2019;9(1)
3. Tsibulskaya E., Lipovka A., Chupakhin A., Dubovoy A., Parshin D., Maslov N. The relationship between the strength characteristics of cerebral aneurysm walls with their status and laser-induced fluorescence data. *Biomedicines*. 2021;9(5):537
4. Tikhvinskii D., Kuianova J., Kislitsin D., Orlov K., Gorbatykh A., Parshin D. Numerical assessment of the risk of abnormal endothelialization for diverter devices: clinical data driven numerical study. *J Pers Med.* 2022;12(4):652
5. Tikhvinskii D.V., Merzhoeva L.R., Chupakhin A.P., Karpenko A.A., Parshin D.V. Computational analysis of the impact of aortic bifurcation geometry to AAA haemodynamics. *Rus J Numer Anal Math Model*. 2022;37(5):311-329
6. Tikhvinsky et al. The role of asymmetry and volume of thrombotic masses in the formation of local deformation of the aneurysmal-altered vascular wall: an in vivo study and mathematical modeling, 2024, PLos One (accepted)

doi 10.18699/bgrs2024-12.1-03

Связь генетического кодирования с циклическими кодами Грея. Проблема наследования биоциклов

Петухов С.В.
Институт машиноведения им. А.А. Благонравова РАН, Москва, Россия
spetoukhov@gmail.com

Ключевые слова: генетический код; алфавиты ДНК; биологические циклы; коды Грея; матрицы; карты Карно; булевы функции; кривая Гильберта; бинарные оппозиции

Мотивация и цель: Живой организм представляет собой огромный хор генетически наследуемых и согласованных циклических процессов. Хрономедицина с древних времен считала, что все болезни являются результатом нарушения этой координации. Напомню несколько примеров циклической природы биотел. Наши белки подчиняются непрерывным циклам жизни и смерти сборки и разборки на аминокислоты. Например, период полураспада гормона инсулина составляет всего несколько минут. Другими словами, генетически унаследованные части нашего тела постоянно циклично умирают и возрождаются. Рассматривая подобные явления, известный физиолог А.Г. Гурвич утверждал: «*Главная проблема биологии – сохранение формы при постоянном обновлении субстрата*» [1]. Эти циклические явления тесно взаимосвязаны с темой внутренних биологических часов, изучение молекулярно-циклических основ которых было отмечено Нобелевской премией в 2017 г.: было показано, что все многоклеточные организмы используют схожий циклический молекулярный механизм внутренних часов для контроля циркадных биоритмов.

Митоз клеток организма происходит циклически. Энергетические затраты на разные циклические события берутся из универсального циклически возобновляемого источника энергии всех биохимических процессов всех живых организмов на Земле: АТФ (аденозинтрифосфорной кислоты). Время жизни одной молекулы АТФ у человека составляет менее одной минуты. В течение суток одна молекула АТФ проходит в среднем 2000–3000 циклов ресинтеза (организм человека синтезирует около 40 кг АТФ в сутки). Другим наглядным примером циклической организации жизни является метаморфоза бабочек, состоящая из стадий бабочка-яйцо-гусеница-куколка-бабочка. Характерно, что бабочку никто не учит тому, как выйти из куколки и начать летать, но она вылезает и начинает летать за счет унаследованных циклических движений крыльев (циклическая генетическая биомеханика). Все необходимое для этого уже имеется в информатике геномного кода. Нейроны и мышечные единицы бабочки работают в соответствии с фундаментальным физиологическим законом «все или ничего», определяющим бинарный принцип их функционирования: нервная клетка или мышечное волокно дают только свои ответы «да» или «нет» под действием различных стимулов по аналогии с булевыми переменными. Эта коллективная деятельность элементов базируется на их взаимной логической координации по аналогии с триггерными ансамблями в компьютере, основанном на логике булевой алгебры. Такие факты привели автора к мысли, что система генетического кодирования способна кодировать наследственные циклические процессы потому, что сама представляет

собой систему циклических кодов, связанную с булевой алгеброй логики. Иными словами, рассматриваемые физиологические процессы цикличны, поскольку генетически закодированы циклическими кодами. Но в современной математике существует большое разнообразие циклических кодов. Целью настоящего исследования стал поиск того типа циклических кодов, который адекватен молекулярным и статистическим особенностям системы генетической информатики. Это может объяснить кодовые основы циклических процессов в живых организмах.

Методы и алгоритмы: Метод исследования основан на учете бинарно-оппозиционных молекулярных особенностей алфавита 4 нуклеотидов ДНК. Эти особенности позволяют представлять алфавитные системы ДНК в форме особых квадрантных матриц. В этих генетических матрицах строки и столбцы нумеруются упорядоченными числами n-битных кодов Грея по аналогии с картами Карно – эффективным методом упрощения выражений булевой алгебры для минимизации числа логических гейтов в логических схемах [2]. Каждой ячейке матрицы присваивается номер в виде кодового слова Грея, представляющего собой конкатенацию номеров соответствующей строки и столбца. Рисунок 1 показывает пример таких матриц для 64 триплетов генетического кода (см. подробности в [2]).

	000 (0)	001 (1)	011 (2)	010 (3)	110 (4)	111 (5)	101 (6)	100 (7)
000 (0)	CCC 000000 (0)	CCA 000001 (1)	CAA 000011 (2)	CAC 000010 (3)	AAC 000110 (4)	AAA 000111 (5)	ACA 000101 (6)	ACC 000100 (7)
001 (1)	CCT 001000 (15)	CCG 001001 (14)	CAG 001011 (13)	CAT 001010 (12)	AAT 001110 (11)	AAG 001111 (10)	ACG 001101 (9)	ACT 001100 (8)
011 (2)	CTT 011000 (16)	CTG 011001 (17)	CGG 011011 (18)	CGT 011010 (19)	AGT 011110 (20)	AGG 011111 (21)	ATG 011101 (22)	ATT 011100 (23)
010 (3)	CTC 010000 (31)	CTA 010001 (30)	CGA 010011 (29)	CGC 010010 (28)	AGC 010110 (27)	AGA 010111 (26)	ATA 010101 (25)	ATC 010100 (24)
110 (4)	TTC 110000 (32)	TTA 110001 (33)	TGA 110011 (34)	TGC 110010 (35)	GGC 110110 (36)	GGA 110111 (37)	GTA 110101 (38)	GTC 110100 (39)
111 (5)	TTT 111000 (47)	TTG 111001 (46)	TGG 111011 (45)	TGT 111010 (44)	GGT 111110 (43)	GGG 111111 (42)	GTG 111101 (41)	GTT 111100 (40)
101 (6)	TCT 101000 (48)	TCG 101001 (49)	TAG 101011 (50)	TAT 101010 (51)	GAT 101110 (52)	GAG 101111 (53)	GCG 101101 (54)	GCT 101100 (55)
100 (7)	TCC 100000 (63)	TCA 100001 (62)	TAA 100011 (61)	TAC 100010 (60)	GAC 100110 (59)	GAA 100111 (58)	GCA 100101 (57)	GCC 100100 (56)

Рис. 1. Расположение 64 триплетов в матрице типа Грея, строки и столбцы которой упорядочены в соответствии с 3-битным кодом Грея. В скобках указаны десятичные эквиваленты кодовых слов 6-битного кода Грея

Кроме того, при учете еще одной известной бинарной оппозиции в системе генетического кода, делящей множество 64 триплетов на два равных подмножества (32 триплета с сильными корнями и 32 триплета со слабыми корнями), генетические матрицы типа показанной на рис. 1 представляются в виде мозаичных бинарных матриц из элементов +1 и –1 (см. [2]). Эти матрицы изучаются алгоритмом их диадо-сдвиговой декомпозиции для установления их алгебро-геометрических свойств. Данными методами и алгоритмами анализируются симметрии в расположении элементов системы генетического кодирования в этих матрицах, условно называемых генетическими матрицами типа Грея. Обнаружение таких симметрий позволяет говорить о связи генетического кодирования с иерархией циклических n-битных кодов Грея.

Результаты: Как кодируемые аминокислоты и стоп-кодоны расположены в показанной матрице из 64 построенных таким образом триплетов? Количество вариантов расположения аминокислот с их повторениями для заполнения всей (8×8)-матрицы огромно: $>>10^{100}$ (для сравнения, в физике время жизни Вселенной оценивается в 10^{17} секунд). Будет ли такое расположение хаотичным или вдруг окажется правильным симметричным? На рис. 2 показан случай митохондриального кода позвоночных, считающегося самым древним и симметричным среди диалектов генетического кода.

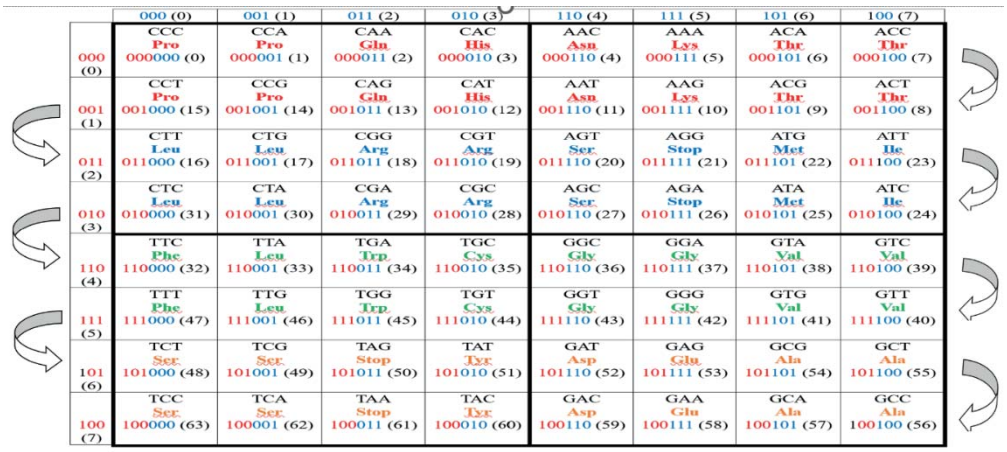


Рис. 2. Расположение 64 триплетов, а также аминокислот и стоп-кодонов, которые они кодируют, в генетической матрице типа Грея (из рис. 1) в случае митохондриального кода позвоночных. Используются общепринятые сокращения аминокислот и стоп-кодонов. Полукруглые скобки по бокам матрицы типа Грея показывают направления порядка кодовых слов 6-битного кода Грея, нумерующих 64 ячейки матрицы и 64 генетических триплета в них

Оказывается, что из океана возможностей природа выбрала симметрично-правильный вариант повторения и расположения аминокислот и стоп-кодонов в матрице типа Грея для 64 триплетов. Показанная матрица кодируемых аминокислот и стоп-кодонов состоит из пар соседних строк с десятичными числами 0-1, 2-3, 4-5, 6-7, одинаковых по составу и расположению аминокислот и стоп-кодонов, показанных цветом. Например, строки 0 и 1 содержат одинаковый по составу и расположению набор красных аминокислот Pro, Gln, His, Asn, Lys, Thr. Строки в каждой из указанных пар 0-1, 2-3, 4-5, 6-7, имеющих в себе повторение аминокислот и стоп-кодонов, отличаются тем, что последовательность 6-битных кодовых слов Грея их 16 ячеек образует циклическую последовательность с единичным расстоянием Хэмминга между соседними ячейками, если читать двоичные числа Грея ячеек в первой из двух строк слева направо, а второй строки – справа налево.

Каждый тип аминокислот в матрице типа Грея (см. рис. 2) сгруппирован таким образом, что занимает только те ячейки, 6-битная нумерация которых образует циклическую последовательность с единичным расстоянием Хэмминга между соседними кодовыми словами. Например, аминокислота Pro находится в клетках с номерами 000000, 000001, 001001, 001000 и т. д. Обнаружено множество других взаимосвязей между кодами Грея и системой генетического кода [2].

Ячейки на картах Карно известны как минтермы, а каждое значение ячейки представляет соответствующее выходное значение логической функции. Эти аналогии между описанными генетическими матрицами типа Грея, имеющими регулярную структурную симметрию, и картами Карно открывают интересные возможности для интерпретации генетических n-плетов как соответствующих выходных значений булевых функций в рамках авторской концепции наследуемой логики генетического интеллекта. Эти исследования связи между системой генетического кодирования и циклическими кодами Грея сопряжены с мыслями основоположника математической логики Г. Буля и его знаменитой книгой «Исследование законов мышления, на которых основаны математические теории логики и вероятности» [3]. Можно добавить, что разработчики цифровой логики широко используют коды Грея для передачи многобитной счетчиковой информации в системах с синхронной логикой, работающих на разных тактовых частотах. В таких случаях считается, что логика работает в разных «регионах часов». Это имеет основополагающее значение для проектирования больших микросхем, работающих на разных тактовых частотах [2].

Коды Грея тесно связаны с фрактальной кривой Гильберта, которая позволяет «дискретизировать» любое пространство, создав в нем удобную систему координат. Это имеет отношение к проблеме того, как генетическая информация, записанная на одномерных нитях ДНК, определяет трехмерную морфологию живых тел [2]. Важно, что пространственная упаковка хроматина в геноме оказывается соответствующей кривой Гильберта, представляющей собой его полимерную фрактальную 3D глобулу [4], свободную от узлов и вынесенную на обложку журнала *Science* (2009, т. 326).

Диадо-сдвиговая декомпозиция упомянутых выше мозаичных генетических матриц с компонентами +1 и –1 обнаруживает, что они являются матричными представлениями гиперкомплексных чисел, связанных со сплит-кватернионами Кокла и дисковой конформной моделью Пуанкаре гиперболической геометрии Лобачевского, имеющей обширные применения в системах искусственного интеллекта [2].

Выводы: Система генетического кодирования структурно связана с иерархией циклических кодов Грея, фрактальной кривой Гильберта, гиперкомплексными числами и геометрией Лобачевского. Эти факты свидетельствуют в пользу того, что она является системой циклического кодирования, определяющей многое в наследовании огромного хора циклических процессов и структур в живых телах. Присущие системе генетического кодирования симметрии обнаруживаются при ее представлении в форме генетических матриц типа Грея, аналогичных картам Карно из теории булевых функций математической логики. Это позволяет интерпретировать генетические n-плеты как выходные значения булевых функций в теме логико-математических основ генетического наследования, а также возвращает к мыслям Дж. Буля в его книге о логико-математических законах мышления [3]. Полученные результаты позволяют развивать новые модельные подходы в сфере алгебраической биологии, генетики, а также генетического и искусственного интеллекта [2, 5].

Relationship between genetic coding and cyclic Gray codes. The problem of inheritance of biocycles

Petoukhov S.V.
Mechanical Engineering Research Institute named after A.A. Blagonravov, RAS, Moscow, Russia
spetoukhov@gmail.com

Key words: genetic code; DNA alphabets; biological cycles; Gray codes; matrices; Karnaugh maps; Boolean functions; Hilbert curve; binary oppositions

Motivation and Aim: A living organism is a huge chorus of genetically inherited and coordinated cyclic processes, the mutual coordination of which is maintained throughout ontogenesis. Since ancient times, chronomedicine has believed that all diseases are the result of a violation of this coordination. Let me remind some examples on cyclic nature of bio-bodies. Our proteins are obeyed to continuous life-death cycles of their assembling and disassembling into amino acids. For example, the half-life of the hormone insulin is only a few minutes. In other words, genetically inherited parts of our body are constantly dying and being reborn in a cyclical manner. Considering such phenomena, the renowned physiologist A.G. Gurvich claimed: "*The main problem in biology is maintaining shape while constantly renewing the substrate*" [1]. These cyclic phenomena is closely interrelated with the topic of internal biological clock, studying of molecular cyclic bases of which was awarded by Nobel prize in 2017: it was shown that all multicellular organisms, including humans, utilize a similar cyclic molecular mechanism of the internal clock to control circadian biorhythms. Somatic cells of the body divide cyclically. The energy costs for these cyclic events are taken from the universal source of energy for all biochemical processes of all living organisms on Earth: ATP (adenosine triphosphoric acid). The lifetime of one ATP molecule in humans is less than one minute. During the day, one ATP molecule goes through an average of 2000–3000 cycles of resynthesis (the human body synthesizes about 40 kg of ATP per day). Another illustrative example of cyclic organization in life is the metamorphosis of butterflies, consisting of the stages butterfly-egg-caterpillar-pupa-butterfly.

The neurons and muscle units of the butterfly work in accordance with the fundamental physiological law of all-or-none, which determines the binary principle of their functioning: a nerve cell or a muscle fiber give only their answers "yes" or "no" under the action of different stimulus by analogy with the Boolean variables. The collective activity of those elements is based on their mutual logical coordination by analogy with the coordination of trigger ensembles in a computer based on Boolean algebra logic. Such facts led the author to the idea that the genetic coding system is capable of encoding inherited cyclic processes because it itself is a system of cyclic codes connected with Boolean algebra of logic. In other words, the physiological processes in question are cyclical because they are genetically encoded by cyclic codes. But in modern mathematics there is a wide variety of types of cyclic codes. The aim of the present study is to search for the those type of cyclic codes that is adequate to the molecular and statistical features of the genetic informatics system. This can explain coded bases of cyclic procecesses in living bodies.

Methods and Algorithms: The research method is based on taking into account the binary oppositional molecular features of the 4 nucleotides of DNA alphabet. These features

make it possible to represent DNA alphabetic systems in the form of special quadrant matrices. In these genetic matrices, the rows and columns are numbered by ordered numbers of n-bit Gray codes, similar to Karnaugh maps, an effective method for simplifying Boolean algebra expressions to minimize the number of logic gates in logic circuits [2]. Each cell of the matrix is assigned a number in the form of a Gray code word, which is a concatenation of the numbers of the corresponding row and column. Figure 1 shows an example of such matrices for 64 triplets of the genetic code (see details in [2]).

	000 (0)	001 (1)	011 (2)	010 (3)	110 (4)	111 (5)	101 (6)	100 (7)
000 (0)	CCC 000000 (0)	CCA 000001 (1)	CAA 000011 (2)	CAC 000010 (3)	AAC 000110 (4)	AAA 000111 (5)	ACA 000101 (6)	ACC 000100 (7)
001 (1)	CCT 001000 (15)	CCG 001001 (14)	CAG 001011 (13)	CAT 001010 (12)	AAT 001110 (11)	AAG 001111 (10)	ACG 001101 (9)	ACT 001100 (8)
011 (2)	CTT 011000 (16)	CTG 011001 (17)	CGG 011011 (18)	CGT 011010 (19)	AGT 011110 (20)	AGG 011111 (21)	ATG 011101 (22)	ATT 011100 (23)
010 (3)	CTC 010000 (31)	CTA 010001 (30)	CGA 010011 (29)	CGC 010010 (28)	AGC 010110 (27)	AGA 010111 (26)	ATA 010101 (25)	ATC 010100 (24)
110 (4)	TTC 110000 (32)	TTA 110001 (33)	TGA 110011 (34)	TGC 110010 (35)	GGC 110110 (36)	GGA 110111 (37)	GTA 110101 (38)	GTC 110100 (39)
111 (5)	TTT 111000 (47)	TTG 111001 (46)	TGG 111011 (45)	TGT 111010 (44)	GGT 111110 (43)	GGG 111111 (42)	GTG 111101 (41)	GTT 111100 (40)
101 (6)	TCT 101000 (48)	TCG 101001 (49)	TAG 101011 (50)	TAT 101010 (51)	GAT 101110 (52)	GAG 101111 (53)	GCG 101101 (54)	GCT 101100 (55)
100 (7)	TCC 100000 (63)	TCA 100001 (62)	TAA 100011 (61)	TAC 100010 (60)	GAC 100110 (59)	GAA 100111 (58)	GCA 100101 (57)	GCC 100100 (56)

Fig. 1. The arrangement of 64 triplets in the Gray-type matrix whose rows and columns are ordered in accordance with 3-bit Gray code. The decimal equivalents of codewords of 6-bit Gray code are indicated in parentheses

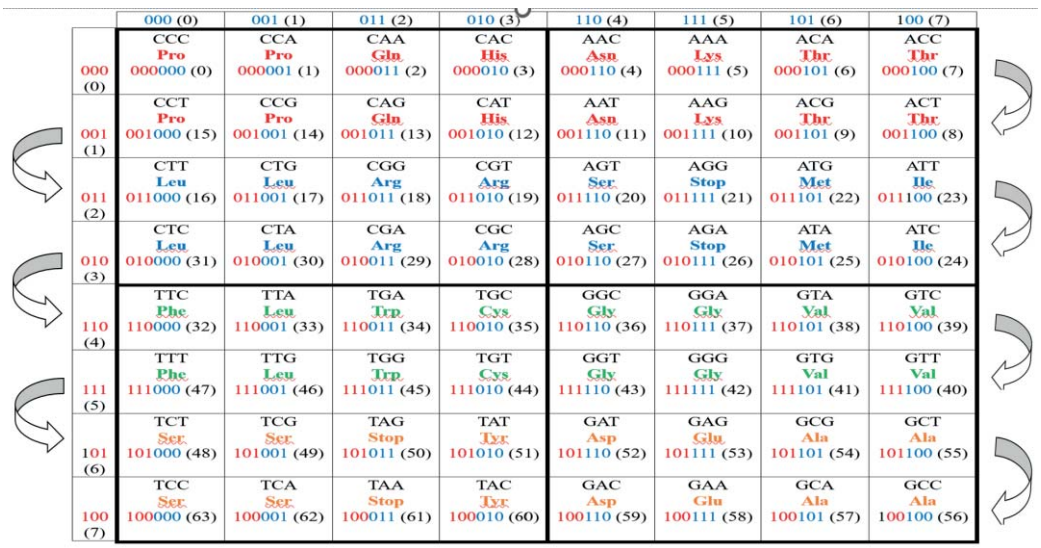


Fig. 2. The location of 64 triplets and the amino acids and stop-codons they encode in the Gray-type matrix (from Fig. 1) in the case of the Vertebrate Mitochondrial Code. Common abbreviations for amino acids and stop codons are used. Semicircular brackets on the sides of the Gray-type matrix show the directions of the codewords order of the 6-bit Gray code, which numerates 64 matrix cells and their 64 triplets

In addition, taking into account another well-known binary opposition in the genetic code system, dividing the set of 64 triplets into two equal subsets (32 triplets with strong roots and 32 triplets with weak roots), genetic matrices of the type shown in Figure are represented in the form of mosaic binary matrices of elements +1 and –1 (see [2]). These matrices are studied by their dyadic-shift decomposition algorithm to establish their algebraic-geometric properties. These methods and algorithms analyze symmetries in the arrangement of elements of the genetic coding system in these matrices, conventionally called Gray-type genetic matrices. The discovery of such symmetries allows us to talk about the connection between genetic coding and the hierarchy of cyclic n-bit Gray codes.

Results: How are the encoded amino acids and stop codons arranged in the shown matrix of 64 triplets constructed in this way? The number of options for the arrangement of amino acids with their repetitions to fill the entire (8×8) matrix is huge: $>>10^{100}$ (for comparison, in physics the lifetime of the Universe is estimated at 10^{17} seconds). Will this arrangement be chaotic or will it suddenly turn out to be correctly symmetrical? Figure 2 shows the case of the Vertebrate Mitochondrial Code, which is considered the most ancient and symmetrical among the genetic code dialects.

It turns out that from the ocean of possibilities, nature chose a symmetrically regular variant of the repetition and arrangement of amino acids and stop-codons in the Gray-type matrix of 64 triplets. The shown matrix of encoded amino acids and stop-codons consists of pairs of adjacent rows with decimal numbers 0-1, 2-3, 4-5, 6-7, identical in composition and arrangement of amino acids and stop-codons, shown in color. For example, lines 0 and 1 contain the same set of red amino acids Pro, Gln, His, Asn, Lys, Thr in composition and location. The lines in each of the indicated pairs 0-1, 2-3, 4-5, 6-7, having the repetition of amino acids and stop-codons in them, differ in that the sequence of 6-bit numbers of their 16 cells forms a cyclic sequence with unit Hamming distance between adjacent cells, if you read the binary Gray numbers of the cells in the first of the two rows from left to right, and the cell numbers of the second row reversely from right to left.Each type of amino acids in the Gray-type matrix (Fig. 2) is grouped in such a way that it occupies only those cells whose 6-bit numbering forms a cyclic sequence with a unit Hamming distance between adjacent codewords. For example, the amino acid Pro is located in cells numbered 000000, 000001, 001001, 001000; and so on. Many other interconnections between Gray codes and the genetic code system are additionally discovered.

Cells in Karnaugh maps are known as minterms, while each cell value represents the corresponding output value of the Boolean function. These analogies between the described genetic Gray-type matrices, having regular structural symmetries, and Karnaugh maps open interesting opportunities to interpret genetic n-plets (duplets, triplets, etc.) as appropriate output values of Boolean functions within the framework of the author's concept of inherited logic of genetic intelligence. This corresponds in some degree to the thoughts of the founder of mathematical logic, G. Boole in his famous book on mathematical-logical bases of thoughts [3]. One can add that digital logic designers widely use Gray codes to transfer multi-bit counter information in systems with synchronous logic that operate at different clock frequencies. In such cases, logic is considered to operate in different “clock regions”. This is fundamental to the design of large chips that operate at many different clock speeds (see detail in [2]).

Gray codes are closely related to the fractal Hilbert curve, which allows you to “discretize” any space, creating a convenient coordinate system in it. This relates to the

problem of how genetic information recorded on one-dimensional DNA strands determines the three-dimensional morphology of living bodies [2]. It is important that the spatial packaging of chromatin in the genome turns out to correspond to the Hilbert curve, which is its polymer fractal 3D globule [4], free of nodes and featured on the cover of the journal *Science* (2009, Vol. 326).

Dyadic-shift decomposition of the above-mentioned mosaic genetic matrices with entries +1 and –1 reveals that they are matrix representations of hypercomplex numbers associated with Cockle split-quaternions and the disk conformal Poincaré model of Lobachevsky hyperbolic geometry, which has extensive applications in artificial intelligence systems [2].

Conclusion: The genetic coding system is structurally related to the hierarchy of cyclic Gray codes, the fractal Hilbert curve, hypercomplex numbers and Lobachevsky geometry. These facts indicate that it is a cyclic character of coding system that determines much of the inheritance of a huge chorus of cyclic processes and structures in living bodies. The symmetries inherent in the genetic coding system are revealed when it is represented in the form of genetic Gray-type matrices, similar to Karnaugh maps from the theory of Boolean functions of mathematical logic. This allows us to interpret genetic n-plets as output values of Boolean functions in the topic of logical-mathematical foundations of genetic inheritance, and also returns to the thoughts of J. Boole in his book on logical-mathematical laws of thinking [3]. The results obtained allow the development of new model approaches in the field of algebraic biology, genetics, as well as genetic and artificial intelligence [2, 5].

Список литературы/References

1. Gurvich A.G. Selected works. Moscow: Medicine, 1977
2. Petoukhov S. Cyclic Gray Codes in Modeling Inherited Cyclic Biostructures and Analysis of Statistical Rules of Genomic DNAs. *Preprints.* 2024;2024020713. doi 10.20944/preprints202402.0713.v1
3. Boole G. An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities. Cambridge University Press, 2009
4. Lieberman-Aiden E., van Berkum N.L., Williams L. et al. Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science*. 2009;326:289-293
5. Petoukhov S.V., He M. Algebraic biology, matrix genetics, and genetic intelligence. Singapore: World Scientific, 2023. doi 10.1142/13468

doi 10.18699/bgrs2024-12.1-04

Интеллектуальные цифровые двойники с системами поддержки принятия решений экспериментальной станции 1-1 «Микрофокус» ЦКП «СКИФ»

Ракшун Я.В.[1]*, Свириденко Д.И.[2], Скляров А.Н.[2]

[1] *Сибирский государственный университет телекоммуникаций и информатики, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

* *rakshun@mail.ru*

Ключевые слова: интеллектуальный цифровой двойник; концептуальное проектирование; системы поддержки принятия решений

Обсуждается проблема создания виртуальных объектов и систем, по своим реакциям на внешнее воздействие максимально соответствующих реальным объектам и системам экспериментальной станции 1-1 «Микрофокус» проекта ЦКП «СКИФ». Рассматриваются вопросы:

- концептуального проектирования объектов и систем станции, включая решение проблем моделирования и модельных расчетов, а также проблемы комбинации модельных расчетов с методами решения обратных задач;
- предиктивной аналитики на основе регламентных и генеративных моделей.

Цифровой двойник (ЦД) экспериментальной станции целесообразно рассматривать с точки зрения оптимизации жизненного цикла научного исследования. Фазы жизненного цикла исследования рассматриваются согласно [1]:

1) Предначальная фаза: оценка острых проблем, выбор тематики исследования;
2) Фаза проектирования: концептуальное проектирование, построение программы исследования;
3) Технологическая фаза: технологическая подготовка, проведение теоретических и/или эмпирических исследований;
4) Рефлексивная фаза: публикации, обсуждения, оценка результатов.

Первая и последняя фазы рассматриваются нами сугубо, как прерогатива человека, при этом внедрение ЦД во вторую и третью фазу исследований способно принести определённые результаты.

Создание ЦД на этапе проектирования станции позволит:

- описать сценарии экспериментов;
- изначально в архитектуре автоматизированной системы управления (АСУ) предусмотреть механизмы и инструменты для подключения всех необходимых цифровых модулей;
- связать воедино инженерную, управляющую, расчетную, визуальную и предиктивные системы станции;

- предварительно проверить принципиальную выполнимость эксперимента, смоделировать результат, а в последующем – программу исследований.

Кроме того, проведение предварительных экспериментов *in silico* с помощью ЦД позволит оценивать время, которое потребуется на проведение исследований.
Оптимизация технологической фазы заключается в непосредственном участии ЦД станции в эксперименте, а именно в:

- оптимизации оптической схемы станции;
- подборе параметров эксперимента;
- принятии решений в процессе эксперимента.

Требования к ЦД экспериментальной станции:
1) ЦД должен включать в себя следующие структурные элементы: цифровая модель станции; цифровая модель образца; система поддержки (принятия) решений (СП(П)Р); набор критериев обоснованности решения; модуль для решения обратных задач;
2) Цифровая модель станции должна актуализироваться на основе реальных показаний приборов;
3) Цифровая модель образца (исследуемого объекта) должна актуализироваться на основе результата решения обратной задачи;
4) СП(П)Р должна взаимодействовать с реальной АСУ станции, опираться в процессе принятия решения на цифровые модели станции и образца, быть ограниченной критериями обоснованности;
5) Критерии обоснованности должны определять целесообразность действий с учетом того, что установка не является идеальной, модель не полна и не точна;
6) Модуль решения обратных задач должен выбран таким образом, чтобы было возможно произвести анализ за требуемое время с определенной степенью достоверности.

Структура ЦД, представленная на рис. 1, с одной стороны, удовлетворяет изложенным требованиям, с другой – позволяет на основе АСУ интегрировать все необходимые цифровые модули. Реальная АСУ станции состоит из Tango сервера [2]; Sardana сервера [3]; модулей Sardana. Tango сервер позволяет управлять Sardana сервером, взаимодействовать с клиентом, а также взаимодействовать с другими Tango серверами. Sardana сервер позволяет выстроить из набора управляемых элементов структуру станции. Модули Sardana, определяют взаимодействие с контроллерами реальных устройств. Ocelot – стандартный пакет для расчета когерентных и частично-когерентных пучков СИ [4].

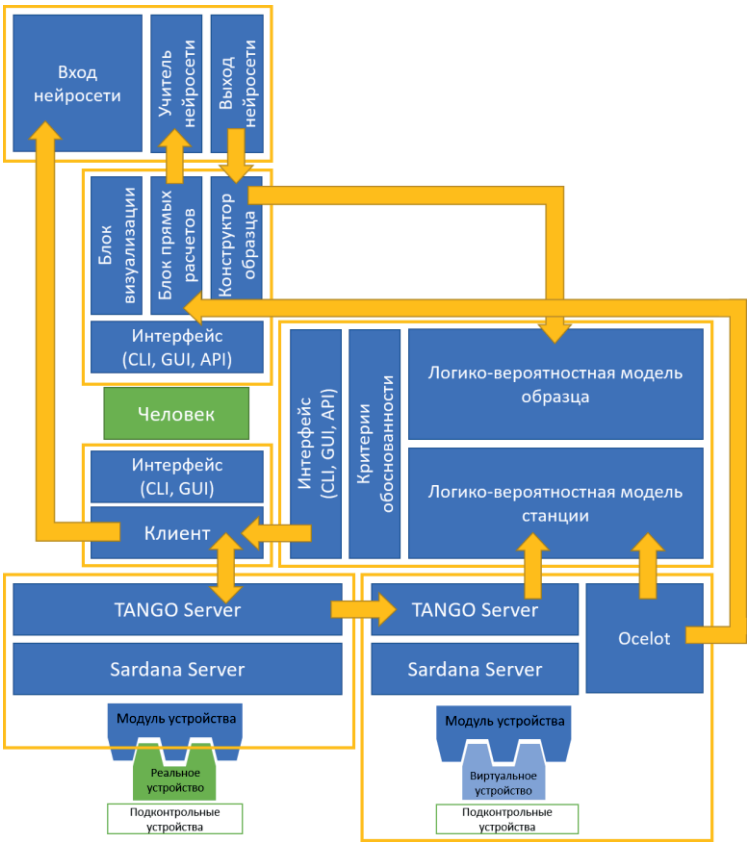


Рис. 1. Блочная архитектура ЦД Экспериментальной станции. Блоки выделены оранжевым контуром, стрелки отражают потоки информации между блоками. Сверху блок решения обратных задач; ниже – цифровая модель образца; справа по центру – СПР; ниже СПР – цифровая модель станции; левее цифрой модели станции – реальная АСУ станции и реальные устройства; над реальной АСУ – модуль для взаимодействия с АСУ станции

Список литературы

1. Кудрявцева Е. А. Проектирование жизненного цикла исследовательского процесса в науке на основе трудов профессоров РГУ им. АН Косыгина. *Кронос*. 2019;10(37):29-32
2. Götz A. et al. TANGO a CORBA based Control System. ICALEPCS2003, Gyeongju, October. 2003;94-98
3. Coutinho T. et al. SARDANA: The software for building SCADAS in Scientific Environments //Proceedings of ICALEPCS2011, Grenoble, WEAAUST01. 2011;607-609
4. Agapov I. et al. OCELOT: A software framework for synchrotron light source and FEL studies. In: Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2014;(768):151-156

doi 10.18699/bgrs2024-12.1-05

Non-uniqueness of cycles in multistage gene networks models

Ayupova N.B., Volokitin E.P., Golubyatnikov V.P.*
Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia
** vladimir.golubyatnikov1@fulbrightmail.org*

Key words: gene network model; negative feedback; cycles; nonlinear dynamical systems

Motivation and Aim: Conditions of existence and stability of periodic trajectories in various piecewise linear models of circular gene networks functioning were described in [1–3]. At the same time, it was shown that in dimensions $n \geq 5$ these models can possess several cycles, see [2, 4]. Here, we construct a similar 3-dimensional gene network model regulated by negative multistep feedbacks which has at least 3 stable cycles. We describe localization of these cycles in the phase portrait of the corresponding dynamical system.

Methods and Algorithms: Consider 3-dimensional piecewise linear dynamical system

$$\frac{dx_j}{dt} = L(x_{j-1}) - x_j; \text{ here and below, } j=1,2,3;\ j-1=3 \text{ for } j=1. \quad (1)$$

The function L is monotonically decreasing, it describes negative feedbacks in this circular gene network. Following [5, 6], where generalized threshold functions were studied in modelling of various natural gene networks, we define this function in the following five-steps way:

$L(w) = 2c$ for $0 \leq w < c-\varepsilon$; $L(w) = c+\varepsilon$ for $c-\varepsilon \leq w < c-\delta$;
$L(w) = c+\delta$ for $c-\delta \leq w < c$; $L(w) = c-\delta$ for $c \leq w < c+\delta$;
$L(w) = c-\varepsilon$ for $c+\delta \leq w < c+\varepsilon$; $L(w) = 0$ for $c+\varepsilon \leq w$.

Here, $c > \varepsilon > \delta > 0$. As in [1, 3], the cube $Q = [0,2c] \times [0,2c] \times [0,2c]$ in the octant $x_j \geq 0$ is positively invariant, i. e., trajectories of its points do not leave it as $t \to \infty$. Let us decompose Q by 15 planes $x_j = c-\varepsilon$, $x_j = c-\delta$, $x_j = c$, $x_j = c+\varepsilon$, $x_j = c+\delta$ to 216 blocks. In each of these blocks, the system (1) splits to independent linear differential equations. We enumerate these blocks by 6-symbols multi-indices $\{s_1 s_2 s_3\}$, where $s_j \in \{\alpha, A, 0, 1, B, \beta\}$ as follows: If $0 \leq x_j < c-\varepsilon$ in this block, then $s_j = \alpha$; for $c-\varepsilon \leq x_j < c-\delta$, we set $s_j = A$; if $c-\delta \leq x_j < c$ then $s_j = 0$; if $c \leq x_j < c+\delta$ then $s_j = 1$; if $c+\delta \leq x_j < c+\varepsilon$ then $s_j = B$; and if $c+\varepsilon \leq x_j$ then $s_j = \beta$.

We need this combinatorics for description, and control of trajectories of the system (1), as well as in corresponding numerical experiments. Similar enumeration was used in [1, 3, 7] in the case of one-step functions L.

Results: The cubes $\hat{Q} = [c-\varepsilon, c+\varepsilon] \times [c-\varepsilon, c+\varepsilon] \times [c-\varepsilon, c+\varepsilon] \subset Q$ and $\tilde{Q} = [c-\delta, c+\delta] \times [c-\delta, c+\delta] \times [c-\delta, c+\delta] \subset \hat{Q}$ are positively invariant domains; this can be shown as for the cube Q. Denote by $W_1 \subset \tilde{Q}$ the union of six blocks listed in the circular diagram

$$\ldots \to \{011\} \to \{010\} \to \{110\} \to \{100\} \to \{101\} \to \{001\} \to \{011\} \to \ldots \quad (2)$$

Let $W_2 \subset \hat{Q} \setminus \tilde{Q}$ be the union of 18 blocks listed in the similar diagram

$$\ldots \to \{BBA\} \to \{B1A\} \to \{B0A\} \to \{BAA\} \to \{BA0\} \to \{BA1\} \to \{BAB\} \to \{1AB\} \to \{0AB\} \to \{AAB\} \to \{A0B\} \to \{A1B\} \to \{ABB\} \to \{AB1\} \to \{AB0\} \to \{ABA\} \to \{0BA\} \to \{1BA\} \to \{BBA\} \to \cdots \quad (3)$$

Each of the blocks of the diagram (3) has either a common vertex, or a common edge with one of the blocks of the diagram (2).

Theorem. 1) The domain W_1 is invariant; it contains exactly one cycle C_1 of the system (1); this cycle is stable and travels from block to block according to the arrows of the diagram (2).

2) If $c \geq 4\varepsilon$ then the domain W_2 contains at least one cycle C_2 of the system (1), this cycle travels from block to block according to the arrows of the diagram (3).

3) If $\varepsilon \geq 4\delta$ then the domain $Q \backslash \hat{Q}$ contains at least one cycle C_3 of the system (1).

4) The cycles $\mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3$ are symmetric with respect to the cyclic permutations of the variables x_j. The system (1) does not have asymmetric cycles.

The cycle C_3 travels from block to block according arrows of one bulky circular diagram analogous to (2), (3). The multi-indices which numerate all 30 blocks of this diagram contain symbols α and β, so all these 30 blocks are incident to the boundary of the invariant domain Q.

Figures 1a and 1b show results of some numerical experiments with the system (1) for two collections of its parameters indicated there. In these Figures, one can see all three cycles of the system (1) described in the Theorem. Here, the cycle C_1 is the smallest one, and C_3 is the largest.

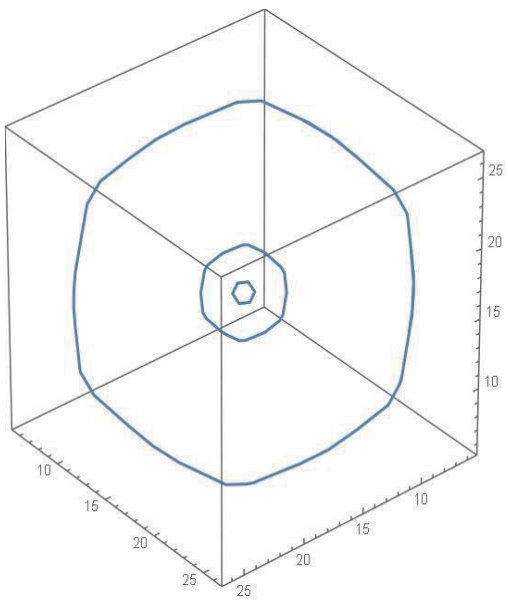


Fig. 1a. Cycles of the system (1); $c = 16; \varepsilon = 4; \delta = 1$

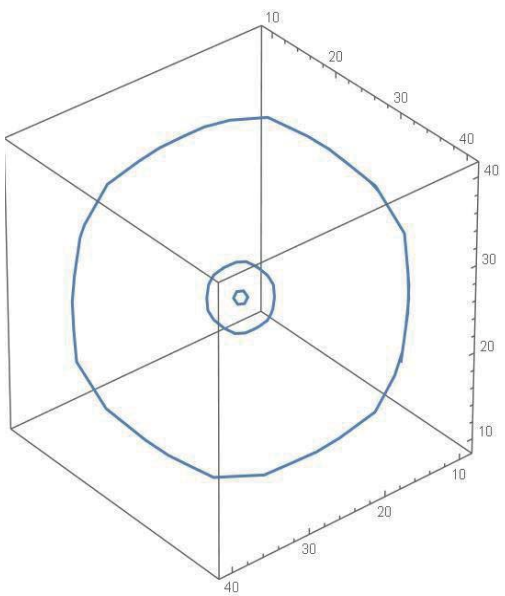


Fig. 1b. Cycles of the system (1); $c = 25; \varepsilon = 5; \delta = 1$

Conclusion and Future Work: For nonlinear dynamical systems of the type (1) with one-step functions in their equations considered in [1–4], non-uniqueness of cycles was observed previously in higher-dimensional cases only, see [2, 4]. Using Generalized Threshold Functions approach described in [5], we have constructed here a 3D gene network model which has three cycles. The cycle C_1 is stable, and the system (1) does not have other cycles in the domain W_1, as in [3]. The cycles C_2 and C_3 give examples

of non-local oscillations, they do not have the bifurcation nature, see [8]. They also should be stable, otherwise we could not see them in the numerical experiments. Similar example of 3D gene network model with two cycles is described in [9].

For these systems of the type (1), we plan to describe invariant surfaces which separate attraction domains of their cycles.

Funding: This work was carried out within the framework of the state task for Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences, projects No. FWNF-2022-0009 and FWNF-2022-0005.

References

1. Akinshin A.A., Golubyatnikov V.P., Golubyatnikov I.V. On some multidimensional models of gene network functioning. *J Appl Ind Math.* 2013;7(3):296-301
2. Likhoshvai V.A., Golubyatnikov V.P., Khlebodarova T.M. Limit cycles in models of circular gene networks regulated by negative feedback loops. *BMC Bioinformatics*. 2020;21(Suppl. 11):255. doi 10.1186/s12859-020-03598-z
3. Ivanov V.V. Attracting limit cycle of an odd-dimensional circular gene network model. *J Appl Ind Math*. 2022;16(3):409-415. doi 10.1134/S199047892203005X
4. Golubyatnikov V.P., Gradov V.S. Non-uniqueness of cycles in piecewise-linear models of circular gene networks. *Sib Adv Math.* 2021;31(1):1-12. doi 10.3103/S1055134421010016
5. Tchuraev R.N., Ratner V.A. A continuous approach with threshold characteristics for simulation of gene expression. In: Molecular Genetic Information Systems. Modelling and Simulation. Berlin: Verlag, 1983;64-80
6. Tchuraev R.N., Galimzyanov A.V. Modeling of actual eukaryotic control gene subnetworks based on the method of generalized threshold models. *Mol Biol*. 2001;35(6):933-939
7. Glass L., Pasternack J.C. Stable oscillations in mathematical models of biological control systems. *J Math Biol*. 1978;6(3):207-223
8. Pliss V.A. Nonlocal Problems in the Theory of Oscillations. Academic Press, 1966
9. Golubyatnikov V.P. On non-uniqueness of cycles in 3D models of circular gene networks. *Chelyabinsk Phys Math J*. 2024;9(1):23-34. doi 10.47475/2500-0101-2024-9-1-23-34

doi 10.18699/bgrs2024-12.1-06

Hidden attractors in models of molecular repressilator

Bondarenko N.E., Glubokikh A.V., Golubyatnikov V.P.*
Novosibirsk State University, Novosibirsk, Russia
** v.golubiatnikov@g.nsu.ru*

Key words: gene network models; step functions; hidden attractor; limit cycle

Motivation and Aim: Detection and description of hidden attractors in phase portraits of dynamical systems is an important problem in applications of nonlinear dynamics, see [1, 2]. This allows to predict behavior of trajectories of these systems, as it was done in [3, 4]. We study here one such system as a model of a simple molecular repressilator. Analogous systems appear in description of various biological processes, such as circadian rhythms and cell cycles in living organisms. The questions of existence and (non)uniqueness of the cycles and description of hidden attractors in these systems are quite necessary, since results of such analysis allow to predict unexpected and potentially dangerous responses of the system to fluctuations.
In this work, a phase portrait of system (1) is constructed. The conditions of compatibility of a hidden attractor and a cycle in this portrait are established.

Methods and Algorithms: We consider a non-linear dynamical system

$$\frac{dx_1}{dt} = L(x_3) - x_1,\ \frac{dx_2}{dt} = L(x_1) - x_2,\ \frac{dx_3}{dt} = L(x_2) - x_3, \qquad (1)$$

as a model of three-components gene network. The variables $x_j(t)$, denote the concentrations of these components, here and below *j*=1,2,3. As in [3,4], monotonically decreasing function *L* describes negative feedbacks in the gene network. Following [5] where multistep feedbacks in gene networks were studied, we define *L* by

$$L(w) = 2c \text{ for } 0 \le w < c - \varepsilon;\ L(w) = c \text{ for } c - \varepsilon \le w < c + \varepsilon;$$

$$L(w) = 0 \text{ for } c + \varepsilon \le w;\ c > \varepsilon > 0.$$

Trajectories of all points of $Q := \{0 \le x_1 \le 2a;\ 0 \le x_2 \le 2a; 0 \le x_3 \le 2a\}$ do not leave this cube as $t \to \infty$. The planes $x_j = a - \varepsilon,\ x_j = a + \varepsilon$ decompose this cube Q to 27 blocks. We describe these blocks by 3-symbols multi-indices $\{s_1 s_2 s_3\}$ as follows

$s_j = 0$ if $0 \le x_j < c - \varepsilon$ in this block; $s_j = 1$ if $c - \varepsilon \le x_j < c + \varepsilon$ in this block; $s_j = 2$ if $c + \varepsilon \le x_j \le 2c$ in this block. Analogous combinatorial description of these blocks in different cases of one-step functions *L* was used in [3,4,6]. It was shown in [3] that all the systems of the type (1) decompose to three independent linear differential equations in each block of similar decompositions. For example, in the block {111} the system (1) has the form

$$\frac{dx_1}{dt} = c - x_1,\ \frac{dx_2}{dt} = c - x_2,\ \frac{dx_3}{dt} = c - x_3.$$

Results: The point S_0 with coordinates $x_1 = x_2 = x_3 = c$ in the block $\{111\}$ is a stable equilibrium point of the system (1). This system does not have other equilibrium points. The system (1) is symmetric with respect to cyclic permutation of the variables $\sigma: x_1 \to x_2 \to x_3 \to x_1$.

Let $W \subset Q$ be the union of 12 blocks listed in the circular diagram

$$\ldots \to \{220\} \to \{210\} \to \{200\} \to \{201\} \to \{202\} \to \{102\} \to \{002\} \to$$
$$\to \{012\} \to \{022\} \to \{021\} \to \{020\} \to \{120\} \to \{220\} \to \cdots \qquad (2)$$

In contrast with Q, this union W is not an invariant domain of the system (1). Trajectories of the point of the face $F_0 = \{120\} \cap \{220\}$ which are sufficiently far from its edge $x_3 = c - \varepsilon$ arrive eventually to the face $F_1 = \{220\} \cap \{210\}$. In the same way, trajectories of F_1 arrive to $F_2 = \{210\} \cap \{200\}$, then to $F_3 = \{200\} \cap \{201\}$, and later, step by step, to all faces F_m which separate incident blocks in the diagram (3), $m = 0,1,\ldots,11$; $F_{12} = F_0$. The analytic representation of these trajectories implies the following

Theorem. If $c > 3\varepsilon$ then the domain W contains at least one cycle C of the system (1). This cycle travels through the blocks of the diagram (2) according to its arrows, and is symmetric with respect to the permutation σ. The system (1) does not have cycles which are not symmetric with respect to σ.

Numerical experiments carried out in the Python programming language, illustrate this theorem. The Fig. 1a and 1b show results of some numerical experiments with the system (1). On Fig. 1a, the parameters of this system satisfy the condition $c > 3\varepsilon$ of the Theorem, and here we see the cycle C, this is an example of nonlocal oscillations in the phase portrait of the system (1) since its equilibrium point S_0 is stable, see [6]. Actually, this point is a Hidden Attractor of the system (1).

On Fig. 1b this condition is not satisfied; here, the trajectories of this system are attracted eventually by the stable equilibrium point S_0 which is depicted in the centers of both Fig. 1.

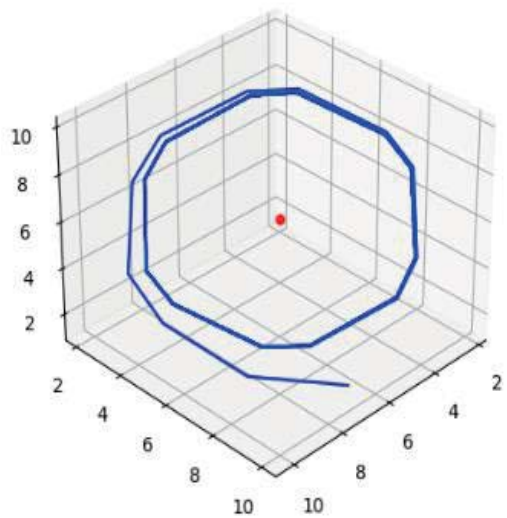


Fig 1a. Limit cycle, $\mathrm{c} = 6$; $\varepsilon = 1.2$

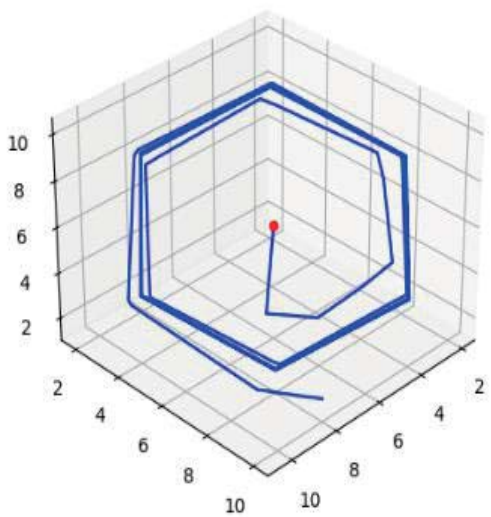


Fig 1b. No cycles, $\mathrm{c} = 6$; $\varepsilon = 2.0001$

Conclusion: The cycle C is a typical example of nonlocal oscillations in gene network models, see [7]. In the phase portrait of the system (1) its unique equilibrium point S_0 is a hidden attractor, see [1], and the cycle C does not appear as a result of bifurcation and parameters variations.

Funding: The study is supported by RSCF (No. 23-21-00019), https://rscf.ru/project/23-21-00019/.

References

1. Dudkowski D., Prasad A., Kapitaniak T. Perpetual points and hidden attractors in dynamical systems. *Phys Lett A*. 2015;379(40-41):2591-2596. doi 10.1016/j.physleta.2015.06.002
2. Lakhova T.N., Kazantsev F.V., Lashin S.A., Matushkin Yu.G. The finding and researching algorithm for potentially oscillating enzymatic systems. *Vavilov J Genet Breed*. 2021;25(3):318-330. doi 10.18699/vj21.035
3. Ayupova N.B., Golubyatnikov V.P. On the uniqueness of a cycle in an asymmetric three-dimensional model of a molecular repressilator. *J Appl Ind Math*. 2014;8(2):153-157. doi 10.1134/S199047891402001X
4. Akinshin A.A., Ayupova N.B., Golubyatnikov V.P., Minushkina L.S. Stratifications and foliations in phase portraits of gene network models. *Vavilov J Genet Breed*. 2022;26(8):758-764. doi 10.18699/VJGB-22-91
5. Galimzyanov A.V., Tchuraev R.N. Dynamic mechanism of phase variation in bacteria based on multistable gene regulatory networks. *J Theor Biol*. 2022;549:111212. doi 10.1016/j.jtbi.2022.111212
6. Golubyatnikov V.P., Ivanov V.V. Cycles in the odd-dimensional models of circular gene networks. *J Appl Ind Math.* 2018;12(4):648-657. doi 10.1134/S1990478918040051
7. Pliss V.A. Nonlocal Problems in the Theory of Oscillations. NY: Academic Press, 1966

doi 10.18699/bgrs2024-12.1-07

Analytical and numerical modeling of a pluripotency gene network

Golubyatnikov V.P.*, Tatarinova E.A.
Novosibirsk State University, Novosibirsk, Russia
** v.golubiatnikov@g.nsu.ru*

Key words: gene network models; positive and negative feedbacks; equilibrium points; stability

Motivation and Aim: We study two 10-dimensional systems of differential equations proposed in [1] as mathematical models of pluripotency gene network functioning. Its main components are the proteins Oct4, Sox2, and Nanog. A simplified version of one of these systems was considered in [2]. Equations of all these systems depend on 62 parameters which were assumed to be fixed in all numerical experiments described in these publications; their values are listed in the Table 1 in [1], see also [3]. In order to extend results of [1] to more general situations, we consider here all these dynamical systems in the cases when the value of all their parameters can vary. For such variable parameters and coefficients of these dynamical systems, we find conditions of non-uniqueness of their equilibrium points, and formulate a criterion of their stability.
Methods and Algorithms: Using methods of qualitative theory of ordinary differential equations, we have fulfilled parametric analysis of these dynamical systems in order to describe these equilibriums, and stability of these equilibriums. We have obtained also a criterion of existence of periodic trajectories near these points, and we have localized these oscillations.
Results: On the basis of these analytical studies, we have realized series of numerical experiments with these dynamical systems for non-fixed values of their parameters described in [1]. Following [4–6], two special cloud programming complexes were elaborated for numerical experiments with these two mathematical models: https://colab.research.google.com/drive/1BKX9KUrgxhOLPkl_WwbCn1EYppu1lu5K ?usp =sharing https://colab.research.google.com/drive/13-pjMCF2SE-v_PxQ6b5ejMFvKM5nUui4
Here, we start with detection and analysis of equilibrium points of these models. The libraries numpy and math were used here for work with data set, and with mathematical functions, the library Matplotlib was used for visualization of the results. In contrast with the previous studies [1, 2], one can vary here all the parameters of these models within some prescribed intervals. The access to these cloud sources is possible from any google account, the values of the current parameters of the systems, and the equilibrium points are demonstrated here as well.
All calculations are performed in the clouds, and presentation of the results is available in a browser. In preparation of similar numerical experiments, it is useful to give a qualitative description of solutions to corresponding differential equations, in order to get their clear presentation. The elaborated interface allows to assign the values of all 62 parameters of the models.
Figures 1 and 2 show results of some numerical experiments with the system (1) described in [1]. On Figure 1, parameters of this system are chosen so that the system has one unstable equilibrium point and two stable equilibriums, the points 1 and 3. On

the Figure 2, these parameters correspond to the case of unique stable equilibrium. Similar variations of multistability of the equilibriums is observed in mathematical analysis and in numerical experiments with the system (4) where following [2], an additional positive feedback loops between Oct4-Sox2-Nanog factors have been taken into account.

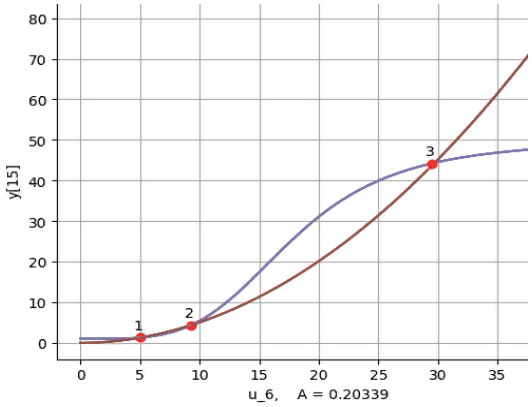


Fig. 1. Three equilibrium points

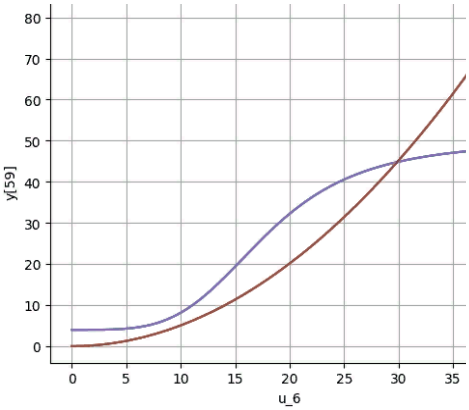


Fig. 2. One equilibrium of the model

Conclusion: The phase portraits of both models under consideration have similar structure; this mathematical observation corresponds to that of [3]: Nanog protein has minimal influence on Oct4 and Sox2 expression.

Funding: The study is supported by RSCF (No. 23-21-00019), https://rscf.ru/project/23-21-00019/.

References

1. Akberdin I.R., Omelyanchuk N.A., Fadeev S.I. et al. Pluripotency gene network dynamics: System views from parametric analysis. *PLoS One*. 2018;13(3):e0194464. doi 10.1371/journal.pone.0194464
2. Chickarmane V., Troein C., Nuber U.A., Sauro H.M., Peterson C. Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol*. 2006;2(9):e123. doi 10.1371/journal.pcbi.0020123
3. Navarro P., Festuccia N., Colby D. et al OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. *EMBO J*. 2012;31(24):4547-4562. doi 10.1038/emboj.2012.321
4. Akinshin A.A., Ayupova N.B., Golubyatnikov V.P. et al. On a numerical model of a circadian oscillator. *Numer Anal Appl*. 2022;15(3):187-196. doi 10.1134/S1995423922030016
5. Bukharina T.A., Akinshin A.A., Golubyatnikov V.P., Furman D.P. Numerical models of the central regulatory circuit of the morphogenesis system of drosophila. *J Appl Ind Math*. 2020;14(2):249-255. doi 10.33048/SIBJIM.2020.23.203
6. Akinshin A.A., Ayupova N.B., Golubyatnikov V.P., Minushkina L.S. Stratifications and foliations in phase portraits of gene network models. *Vavilov J Genet Breed*. 2022;26(8):758-764. doi 10.18699/VJGB-22-91

doi 10.18699/bgrs2024-12.1-08

Software package for analysis of regulatory circuits of bacterial metabolic pathways by mathematical modeling methods

Kazantsev F.V.[1, 2, 3], Lakhova T.N.[1, 2, 3], Khlebodarova T.M.[1, 2], Matushkin Yu.G.[1, 2, 3], Lashin S.A.[1, 2, 3]

[1] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *kazfdr@bionet.nsc.ru*

Key words: mathematical modelling; gene networks; gene regulations

Motivation and Aim: Today there are many tools and libraries available within engineering modeling environments that can be used to create mathematical models of the regulatory circuits of bacterial metabolic pathways. However, such models should be created using approaches and standards recognized by the scientific community. Such approach is the "Elementary Subsystems" approach, on which DSL tools and model representation standards SBML, SBGN, CellML, etc. are based. In the study, we present a working algorithm and its software implementation for reconstructing and analyzing large regulatory circuits, commonly arising in automatic generation problems with dozens/hundreds of subsystems.

Methods and Algorithms: The algorithm is implemented in Python programming language as package of Jupyter Notebooks. The basico library (basico.readthedocs.io) is used to build the model. Computational experiments are performed in the Copasi tool (copasi.org). The final model is implemented in SBML standard (sbml.org).

Results: Nowadays, information on regulatory relationships is built based on automated tools for whole-genome analysis and we obtain dozens/hundreds of nodes in regulatory circuits. The transition in reconstruction from such large-scale regulatory contours to their mathematical models within existing tools providing WYSIWYG approaches (as Copasi, CellDesigner, ECell) of model building is a rather labor intensive process. The developed software package provides a prepared set of steps that automate this transition: preparation of elementary subsystems, their placement in the compartment structure, generation of a combined mathematical model (export in SBML standard) and computational experiments. Setting additional features in the circuit elements annotation is the necessary step of customizing data post-processing methods for summary charts and simulation results visualization.

Funding: The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics SB RAS No. 075-15-2019-1662.

doi 10.18699/bgrs2024-12.1-09

The virtual cell

Kolpakov F.A.*, Akberdin I.R., Kutumova E.O.

Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia

** kolpakov.fa@talantiuspeh.ru*

Key words: virtual cell; mathematical modeling; systems biology; GTRD, BioUML

Motivation and Aim: With the beginning of the 21st century, systems biology, an interdisciplinary science that studies complex interactions in living systems, has received active development. Systems biology involves a specific cycle of research, consisting of theory, analytical or computer modeling to formulate hypotheses about the system, experimental testing of the hypotheses, and then using the obtained data to describe cellular processes, allowing to improve the computer model or refine the theory. Since the ultimate goal is to build a portrait model of interactions in a complex system, experimental techniques used in systems biology must be as detailed as possible. For this purpose, experimental approaches such as transcriptomics, metabolomics, proteomics and other high-throughput omics technologies are used to obtain numerical data. Thus, success in the development of high-throughput automated biomedical technologies has led to a renaissance in modeling biological systems at the molecular-cellular level.

The Virtual Cell project suggests a new approach that can be represented as a pyramid (Fig. 1) where each layer contains less information (10 to 10 000 times) than previous one.

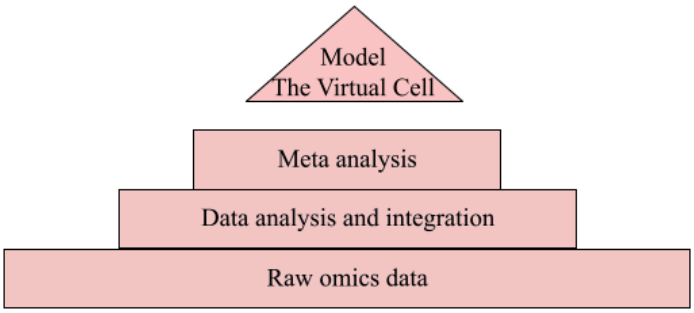


Fig. 1. The virtual cell as a model to aggregate and replace raw omics data

Finally, as the project goal, the constructed virtual cell model should accurately reproduce (replace) the experimental omics data obtained for a given human cell type (cell line):

- gene expression level (transcriptomics);
- protein quantities (proteomics);
- concentration of metabolites (metabolomics).

The developed virtual cell model verified on these omics data will serve as a basis for *in silico* experiments aimed at creating a state-of the-art rational strategy to fundamentally understand and improve the target properties of cells.

Methods and Algorithms: Architecture of The Virtual Cell shown in Figure 2. The key ideas are:

- agent based approach – main cellular processes (transcription, translation, metabolism, etc.) are presented by agents;
- different agents use different approaches to simulate their dynamics. For example, a neutral network is used to predict gene expression level using information from the GTRD [1], KnockTF [2] and FANTOM5 [3]. Atlas of human metabolism provides a genome-scale model of metabolite fluxes [4]. Ordinary differential equations are used to simulate cellular signal transduction pathways and energy metabolism [5]. Corresponding new plugin for BioUML platform [6] is developed;
- there are 3 main shared molecule pools: RNAs, proteins and their complexes and other small molecules and metabolites. They serve as input and output for all agents. Each agent specifies which molecules it accepts as inputs or outputs.
- iterative development of the model allows incrementally refine the model and add new agents;
- there are a number of the virtual cell builds that correspond different cell types (lines) as well as different iterations.

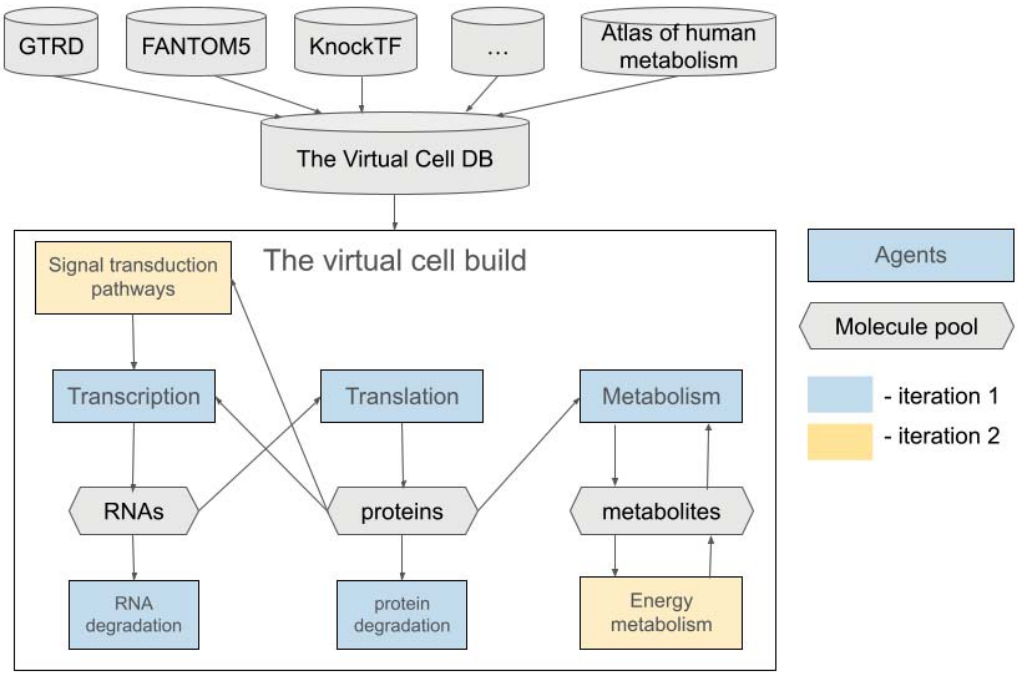


Fig. 2. The Virtual Cell architecture. First two iterations are shown

Conclusion: Here we are suggesting a new iterative approach to build a complicated computational model that should accurately reproduce (replace) the experimental whole genome omics data obtained for a given cell type. The Virtual Cell is an open source project available at https://gitlab.sirius-web.org/virtual-cell.

We look forward to working with other teams to join forces to tackle this challenging task.

Funding: The study was financially supported by the Russian Science Foundation (project No. 24-14-20031, https://www.rscf.ru/en/project/24-14-20031/).

References

1. Kolmykov S., Yevshin I., Kulyashov M. et al. GTRD: an integrated view of transcription regulation. *Nucleic Acids Res.* 2021;49(D1):D104-D111. doi 10.1093/nar/gkaa1057
2. Feng C., Song C., Liu Y. et al. KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors. *Nucleic Acids Res.* 2020;48(D1):D93-D100. doi 10.1093/nar/gkz881
3. FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al. A promoter-level mammalian expression atlas. *Nature.* 2014;507:462-470. doi 10.1038/nature13182
4. Wang H., Robinson J.L., Kocabas P. et al. Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *PNAS.* 2021;118,e2102344118. doi 10.1073/pnas.2102344118
5. Akberdin I.R., Kiselev I.N., Pintus S.S. et al. A Modular Mathematical Model of Exercise-Induced Changes in Metabolism, Signaling, and Gene Expression in Human Skeletal Muscle. *Int J Mol Sci.* 2021;22(19):10353. doi 10.3390/ijms221910353
6. Kolpakov F., Akberdin I., Kiselev I. et al. BioUML-towards a universal research platform. *Nucleic Acids Res.* 2022;50(W1):W124-W131. doi:10.1093/nar/gkac286

doi 10.18699/bgrs2024-12.1-10

Multilevel mathematical model of epileptic seizures

Kondrakhin P.[1]*, Kolpakov F.[1, 2]

[1] *Sirius University of Science and Technology, Sirius, Russia*

[2] *Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia*

* *kondrakhin.py@gmail.com*

Key words: mathematical model; epilepsy; brain network model; biophysical neuron model

Motivation and Aim: There are three approaches to brain modeling: nanoscopic (cellular), focusing on individual neurons; microscopic (population), simulating interactions among neuron populations; and mesoscopic (regional), examining entire brain regions.

The microscopic approach's applicability is limited by the high computational complexity of its models. Mesoscopic models, using methods akin to statistical physics, reduce micro-level complexity to achieve macro-organization. These models simulate neuronal fields instead of individual neurons, generating accurate electroencephalogram (EEG) and local field potential (LFP) signals. However, they often lack biological interpretation, which significantly complicates result validation. Nanoscopic models simulate biophysical processes within individual neurons, providing strong biological significance but failing to reproduce brain signals like EEG, limiting their clinical applicability.

The Virtual Brain (TVB) project, which uses mesoscopic models to forecast surgical outcomes, has made significant progress and is now undergoing clinical trials [1]. However, TVB models do not consider the nanoscopic factors influencing seizures, such as receptor dynamics, gene and metabolic networks, signal transmission pathways within neuron, etc. Our team is developing the BioUML (Biological Universal Modeling Language) platform, which is well-suited for describing the nanoscopic level [2].

Thus, we aim to recreate a series of existing epilepsy models in BioUML that qualitatively reproduce seizure phenomena and then link the mesoscopic and nanoscopic levels, bypassing the microscopic. This will allow the generation of model electrophysiological signals and comparison with experimental recordings, while drug therapy modeling will be conducted at the nanoscopic level.

Methods and Algorithms: We developed a multilevel model using a modular approach in the BioUML platform, decomposing complex systems into modules. This enables the integration of multiple biological system levels within a single model, facilitating the exploration of interactions within subsystems and between different organizational levels.

Numerical calculations used patient-specific connectivity and delay matrices to determine signal transmission strength and time delays due to limited speed of nerve impulse propagation.

The model's equations include stochastic terms to describe natural brain network noise. We integrated a solver for stochastic differential equations based on the Euler–Maruyama method in BioUML.

Results: The multilevel model includes one regional submodel that models interactions between brain regions, with each region corresponds to a cellular submodel calculating

ion concentration dynamics, and a receptor submodel determining the dynamics of AMPA receptor trafficking in dendritic spine (Fig. 1). The receptor submodel regulates synaptic transmission efficiency, while the cellular submodel serves as the excitation source for the corresponding brain region.

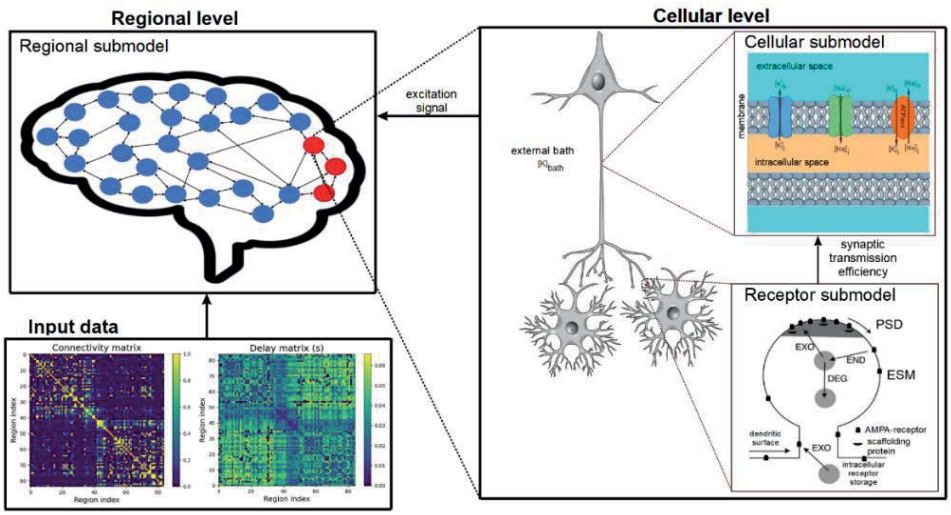


Fig. 1. Multilevel model structure

Simulations were performed for a system of 84 regions, with the epileptic focus set in region 64, corresponding to the right parahippocampal gyrus. All other regions were considered healthy, incapable of autonomous epileptiform activity.

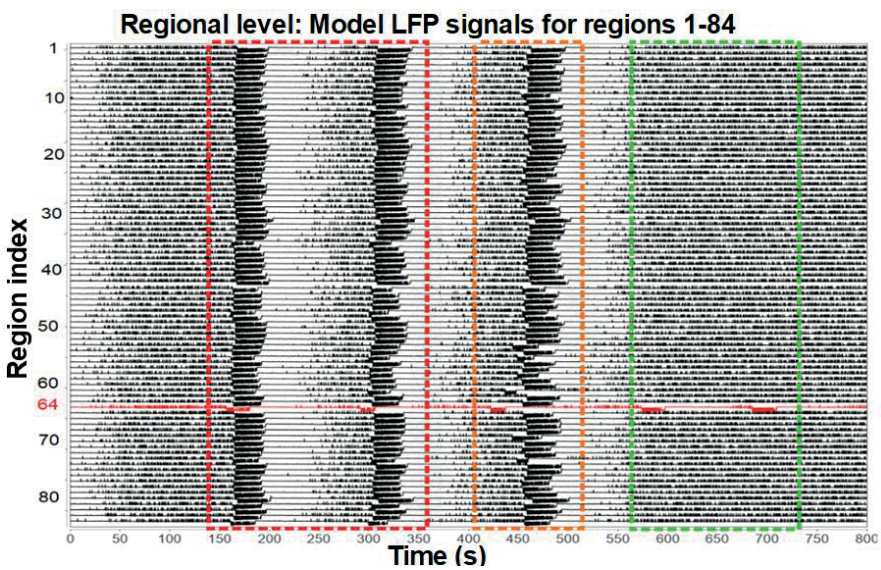


Fig. 2. Dynamics of the regional submodel. The epileptogenic region signal is red, healthy region signals are black. Red dashed frame highlights seizure spread throughout all brain regions, orange dashed frame indicates delayed propagation, and green dashed frame shows localized seizures

Figure 2 depicts the dynamics of model local field potentials for all considered brain regions. The seizure first emerges in the epileptogenic region. It then propagates to

healthy regions, and epileptiform activity is concurrently observed across all brain zones. The system's excitation gradually diminishes, leading to subsequent ictal periods where either the recruitment of healthy regions occurs with a significantly greater delay compared to the initial period, or the seizure is entirely localized in the epileptogenic region.

In Figure 3, the dynamics of the cellular submodel associated with the epileptogenic region is presented. During each seizure event, there is an increase in extracellular potassium concentration, leading to membrane depolarization, and the generation of action potentials by the cell.

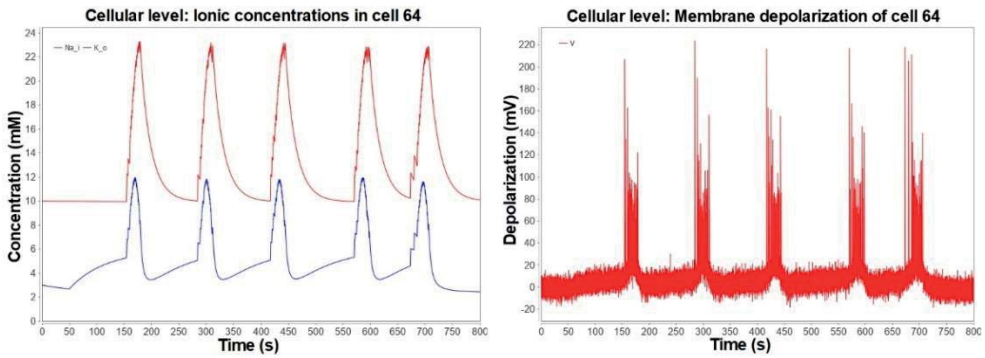


Fig. 3. Dynamics of the cellular submodel. Left: ion concentrations (red – intracellular sodium, blue – extracellular potassium). Right: membrane depolarization

It is demonstrated that the dynamics of individual submodels used in the structure of the multilevel model corresponds to the dynamics of the original models, validated by comparison with experimental data conducted by their authors. Theoretical justification of the connections uniting submodels into a unified structure is provided.

Conclusion: This study presents a novel developed multilevel mathematical model of epileptic seizures, describing interactions between brain regions while considering the biophysical processes occurring within individual neurons.

The constructed model for the first time makes it possible to simulate the dynamics of seizure onset, propagation and termination simultaneously at the cellular and regional levels of brain organization.

Analysis of numerical simulation results of the model demonstrated a qualitative reproduction of the main processes observed during seizures. Thus, the constructed multilevel model can be reasonably used in epilepsy research.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-10-2021-093).

References

1. Sanz L.P., Knock S.A., Woodman M.M., Domide L., Mersmann J., McIntosh A.R., Jirsa V.K. The Virtual Brain: a simulator of primate brain network dynamics. *Front Neuroinform*. 2013;7:10. doi 10.3389/fninf.2013.00010
2. Kolpakov F., Akberdin I., Kiselev I. et al. BioUML – towards a universal research platform. *Nucleic Acids Res*. 2022;50(W1):W124-W131. doi 10.1093/nar/gkac286

doi 10.18699/bgrs2024-12.1-11

On cycles in 6D circular gene networks models

Minushkina L.S.*
Novosibirsk State University, Novosibirsk, Russia
** l.minushkina@g.nsu.ru*

Key words: gene network models; cycles; phase portraits; invariant surfaces; dynamical systems

Motivation and Aim: In this paper we study a six-component gene network model with positive and negative feedback. The aim is to find the conditions of existence of a cycle and construct an invariant surface for trajectories of a dynamical system.
Methods and Algorithms: We consider a non-linear dynamical system

$$\frac{dx_i}{dt} = L_i(y_{i-1}) - \Gamma_i(x_i), \frac{dy_i}{dt} = G_i(x_i) - \gamma_i(y_i) \quad (1)$$

where positive variables x_i, y_i denote the concentrations of components of gene network, L_i are smooth monotonically decreasing functions and Γ_i, G_i, γ_i are smooth monotonically increasing functions. Hereinafter $i = 1, 2, 3$, $i - 1 := 3,\ i = 1$.
A symmetric six-dimensional model of synthetic molecular repressilator was suggested in [1] with linear degradation rate Γ_i, γ_i and synthesis rate L_i, G_i. A similar smooth non-symmetric dynamical system was studied in [2] with smooth L_i and linear Γ_i, γ_i, G_i. In paper [3] the sufficient conditions of existence of a cycle were stated for an analogous six-dimensional system with discontinuous right-hand sides where negative and positive feedback are expressed by decreasing and increasing step functions and degradation rate is linear.
In this paper, we consider only such cases when for all i inequalities

$$\sup \Gamma_i \ \geq\ L_i(0), \qquad \sup \gamma_i \geq \sup G_i$$

hold so the values of $\Gamma_i^{-1}\big(L_i(0)\big)$ and $\gamma_i^{-1}(\sup G_i)$ are defined.
The trajectories of system (1) entering the domain

$$\mathcal{Q} = \prod_{i=1}^{3}\big[0, \Gamma_i^{-1}\big(L_i(0)\big)\big] \times \big[0, \gamma_i^{-1}(\sup G_i)\big] \quad (2)$$

do not leave it as time grows. The domain $\mathcal{Q}$ contains a unique equilibrium point S_0 of system (1) and the coordinates of S_0 are determined from the equation

$$y_3 = \gamma_3^{-1}(G_3(\Gamma_3^{-1}(L_3(\gamma_2^{-1}(G_2(\Gamma_2^{-1}(L_2(\gamma_1^{-1}(G_1(\Gamma_1^{-1}(L_1(y\_3)))))))))))$$

Let us divide the invariant parallelepiped $\mathcal{Q}$ into 64 smaller parallelepipeds (blocks) by drawing the hyperplanes through the point S_0 and enumerate them by the binary multiindices as in [5]. In this paper, we study the trajectories of system (1) passing through the blocks according to the following diagram

$$\ldots\to\{110011\}\to\{010011\}\to\{000011\}\to\{001011\}\to\{001111\}\to\{001101\}\to \quad (3)$$
$$\to\{001100\}\to\{101100\}\to\{111100\}\to\{110100\}\to\{110000\}\to\{110010\}\to\ldots$$

Results: A characteristic polynomial of the linearization matrix in a small neighborhood of S_0 has a form

$$P(\lambda) = (\lambda + p_1)(\lambda + p_2)(\lambda + p_3)(\lambda + q_1)(\lambda + q_2)(\lambda + q_3) + k_1k_2k_3l_1l_2l_3,$$

where $p_i = \Gamma_i', q_i = \gamma_i', l_i = -\mathrm{L}_i', k_i = \mathrm{G}_i'$.

For sufficiently large k_i, l_i compared to p_i, q_i the equilibrium point S_0 is hyperbolic and conditions of the Hartman–Grobman theorem are met so the characteristic polynomial has at least a pair of complex conjugate roots λ_1, λ_2 with positive real part.

Theorem. If the characteristic polynomial of linearization matrix of dynamical system (1) has at least two complex conjugate roots with positive real parts and does not have imaginary roots, then system (1) has a cycle $\mathcal{C}$ passing through the blocks of diagram (2).

According to the Hartman–Grobman theorem, there exist neighborhoods U_1, U_2, V_1, V_2 of hyperbolic equilibrium point S_0 and a homeomorphism h: $U_1 \cup U_2 \longrightarrow V_1 \cup V_2$ such that $\Phi = h^{-1} \circ D\Phi \circ h$ on U_1. The plane Π_1^2 spanned by vectors $Re\ Z$ and $Im\ Z$ where Z is a complex eigenvector corresponding to λ_1 is contained in the union of the blocks of diagram (2). Let us draw a segment in a small neighborhood of S_0 and after a certain number of iterations of the linear part $D\Phi$ of the Poincaré map Φ we obtain a nested segment sequence bounded by the fixed point U of $D\Phi$. With continuous change of coordinates h a given sequence goes into a nested arc sequence bounded by the arc joining S_0 and $h^{-1}(U)$. The trajectories of all points lying on this arc form a two-dimensional invariant surface of system (1) bounded by a cycle $\mathcal{C}$.

Conclusion: Thus, the sufficient conditions for the existence of a cycle in the invariant domain have been proven for circular gene network model with nonlinear degradation. It has been also shown that the cycle limits a two-dimensional invariant surface containing the trajectories of considered dynamical system. The results generalize the results obtained in [2, 3].

Funding: The study is supported by RSCF (No. 23-21-00019) https://rscf.ru/project/23-21-00019/.

References

1. Elowitz M.B., Leibler S. A synthetic oscillatory network of transcriptional regulator. *Nature.* 2000;403:335-338. doi 10.1038/35002125
2. Ayupova N.B., Golubyatnikov V.P., Kazantsev M.V. On existence of a cycle in an asymmetric model of a molecular repressilator. *Numer Anal Appl*. 2017;10(2):101-107. doi 10.1134/S199542391702001X
3. Golubyatnikov V.P., Minushkina L.S. Monotonicity of the Poincaré mapping in some models of circular gene networks. *J Appl Ind Math.* 2019;13(3):472-479. doi 10.1134/S1990478919030086
4. Golubyatnikov V.P., Minushkina L.S. On uniqueness and stability of a cycle in one gene network. *Sib Electr Math Rep.* 2021;18(1):464-473. doi 10.33048/semi.2021.18.032
5. Glass L., Pasternack J.C. Stable oscillations in mathematical models of biological control systems. *J Math Biol.* 1978;6:207-223

12 Симпозиум «Математические проблемы биоинформатики и системной компьютерной биологии. Анализ больших генетических данных и искусственный интеллект»

Symposium “Mathematical problems of bioinformatics and systems computational biology. Big genetic data analysis and artificial intelligence”

doi 10.18699/bgrs2024-12.2-01

Математическое моделирование динамики асептического воспаления: иерархический подход

Воропаева О.Ф.

Федеральный исследовательский центр информационных и вычислительных технологий, Новосибирск, Россия

vorop@ict.nsc.ru

Ключевые слова: асептическое воспаление; иерархия математических моделей; качественные свойства; мультистабильность; вычислительный эксперимент; экспериментально-вычислительный подход; поляризация макрофагов; цитокины; нейрогенные факторы; противовоспалительная терапия

Мотивация и цель: Работа посвящена проблеме математического анализа фундаментальных клеточно-молекулярных механизмов, обеспечивающих реакцию на локальное повреждение при пренебрежимо малом влиянии микробного загрязнения на раневой процесс. Актуальность темы, в которой особое значение имеет формально-математическое описание этих механизмов с привлечением количественных и качественных лабораторных и клинических данных, обусловлена широтой области приложений – от хирургических кожных ран до повреждений миокарда или мозга при инфарктах.

Известно, что воспалительный ответ на повреждение любой этиологии обеспечивается, прежде всего, системой врожденного иммунитета, которая регулируется нервной и эндокринной системами через сложные взаимодействия иммунных клеток между собой, с другими клетками и биоактивными веществами (медиаторами воспаления и нейромедиаторами, цитокинами, гормонами и другими эффекторными молекулами). Воспаление выполняет защитную функцию, но одновременно может стать важным фактором патогенеза, усиливающим повреждение. Управление воспалительной реакцией, основанное на более глубоком понимании ее механизмов, – одно из активно развивающихся направлений современной биомедицины, которое связано с проблематикой раннего прогноза характера иммунного ответа и поиска более эффективных стратегий для снижения уровня повреждения и ускорения восстановительного процесса (обзор литературы можно найти в [1–6]).

Методы и алгоритмы: Механизмы многих заболеваний, важной частью патогенеза которых является воспаление, являются крайне сложными и запутанными. Широта перечня наиболее значимых факторов асептического воспаления, их многофункциональность и пластичность, а также труднодоступность детальных экспериментальных и клинических измерений требуют более гибкого подхода в выборе оптимального набора параметров состояния при математическом моделировании динамики процессов в поврежденных участках ткани. Поэтому возникает потребность в построении иерархии математических моделей разной степени биологической детализации и математической сложности, которые должны обладать некоторым набором заданных биологически обусловленных математических свойств. Необходимым

условием, определяющим адекватность моделей, является также качественное и количественное соответствие экспериментальным данным.

Разработка новых математических моделей производится в соответствии с общепринятой методологией идентификации систем и математического моделирования. В работе используется оригинальная экономичная вычислительная технология идентификации математических моделей, основанная на известном экспериментально-вычислительном подходе и идее декомпозиции–сборки систем дифференциальных уравнений. Численное решение задачи структурной и параметрической идентификации дифференциальных уравнений основано на использовании методов типа предиктор–корректор и А-устойчивых численных методов семейства Адамса, генетического алгоритма и метода расщепления по пространственным направлениям [7].

Результаты: В работе приводятся результаты нескольких циклов исследований в области создания иерархий математических моделей, сформулированных в виде начальных и начально-краевых задач для систем обыкновенных дифференциальных уравнений, уравнений с запаздыванием и реакционно–диффузионных уравнений.

Наиболее лаконичное представление о молекулярно-клеточных механизмах асептического воспаления используется в иерархии математических моделей развития повреждения миокарда при инфаркте II типа. В рамках этой иерархии даны оценки эффективности гипотетических стратегий управления поляризацией и перепрограммированием макрофагов, в том числе с точки зрения «терапевтического окна». Использование иерархического подхода позволило проиллюстрировать достаточную широту возможностей локально-однородного приближения и дать представление о роли пространственных эффектов в развитии воспалительного ответа при инфаркте [2–4].

Разработана новая иерархия математических моделей динамики асептического воспалительного процесса триггерного типа, основанная на модели [5, 6]. Модели этой иерархии характеризуются разной степенью полноты набора параметров состояния, но все они могут быть использованы для предварительной оценки риска прогрессирования острого асептического воспаления в хроническое по начальным данным.

Предложена и исследована иерархия новых математических моделей функционирования сложного механизма клеточно-молекулярной реакции на повреждение, включающей иммунный и нейрогенный ответы в их взаимодействии.

Все разработанные модели успешно прошли валидацию с привлечением экспериментальных данных, исследована их чувствительность к неопределенностям в параметрах, зависимость от начальных условий и область применимости.

Финансирование: Работа выполнена в рамках государственного задания Минобрнауки России для Федерального исследовательского центра информационных и вычислительных технологий.

Mathematical modeling of aseptic inflammation dynamics: a hierarchical approach

Voropaeva O.F.
Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia
vorop@ict.nsc.ru

Key words: aseptic inflammation; hierarchy of mathematical models; qualitative properties; multistability; computational experiment; experimental computational approach; polarization of macrophages; cytokines; neurogenic factors; hypothetical therapeutic strategies

Motivation and Aim: The work is devoted to the problem of mathematical analysis of the fundamental cellular and molecular mechanisms that provide a response to local damage with negligible effect of microbial contamination on the wound healing process. The relevance of the topic is due to the breadth of the field of applications – from surgical skin wounds to myocardial or brain ischemic damage.
It is well known that the inflammatory response to damage of any etiology is provided primarily by the innate immune system, which is regulated by the nervous and endocrine systems through complex interactions of immune cells with each other, other cells and bioactive substances (inflammatory mediators and neurotransmitters, cytokines, hormones and other effector molecules). Inflammation performs a protective function, but at the same time it can become an important factor in pathogenesis, enhancing damage. The management of an inflammatory response based on a deeper understanding of its mechanisms is one of the actively developing areas of modern biomedicine. It is especially relevant in connection with the problems of early prediction of the nature of the immune response and the search for more effective strategies to reduce the level of damage and accelerate the recovery process. A detailed review of the research can be found in [1–6].
Methods and Algorithms: The mechanisms of many diseases, an important part of the pathogenesis of which is inflammation, are extremely complex and confusing. One of the most common variants of inflammation is considered to be the process in the necrosis zone with a local violation of tissue viability. However, the list of the most significant factors of aseptic inflammation is too wide. And these factors are characterized by versatility and plasticity, and their detailed experimental measurements are expensive and too complicated. Because of this, it is necessary to take a more flexible approach when choosing the optimal set of state parameters for mathematical modeling of the dynamics of processes in damaged tissue areas.
Therefore, it is necessary to create a hierarchy of mathematical models with varying levels of biological detail and mathematical complexity. These models must have a specific set of biologically determined mathematical properties. A necessary condition determining the adequacy of models is also qualitative and quantitative compliance with experimental data.
The development of new mathematical models is carried out in accordance with the generally accepted methodology of system identification and mathematical modeling. The work uses an original cost-effective computational technology for identifying mathematical models based on the well-known experimental computational approach and the idea of decomposition-assembly of systems of differential equations. The

numerical solution of the problem of structural and parametric identification of differential equations is based on the use of predictor–corrector type methods and A-stable numerical methods of the Adams family, a genetic algorithm and a spatial splitting method [7].

Results: The report presents the results of several research cycles in the field of creating hierarchies of mathematical models formulated in the form of initial or initial–boundary value problems for nonlinear systems of ordinary differential equations, equations with time delay and reaction–diffusion equations [2–6].

The most concise representation of the molecular and cellular mechanisms of aseptic inflammation is used in the hierarchy of mathematical models of myocardial infarction [2, 3]. The hierarchical approach made it possible to illustrate a sufficient breadth of possibilities for a locally homogeneous approximation and to give an idea of the role of spatial effects in the development of an inflammatory response in heart attack [3]. Within this hierarchy, estimates of the effectiveness of hypothetical strategies for controlling polarization and reprogramming of macrophages are given, including from the perspective of the "therapeutic window" [4].

A new hierarchy of trigger-type models of the dynamics of the aseptic inflammatory process has been developed. The models in this hierarchy are characterized by varying degrees of completeness of the set of state parameters, but they can all be used to pre-assess the risk of progression of acute aseptic inflammation to chronic according to initial data.

A hierarchy of new mathematical models of the functioning of a complex mechanism of cellular and molecular response to damage, including immune and neurogenic responses in their interaction, is proposed and investigated. These models include the trigger-type model [5, 6] as their basis.

All the developed models have been successfully validated using experimental data, their sensitivity, dependence on initial conditions and the scope of applicability have been investigated.

Funding: The research was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation for Federal Research Center for Information and Computational Technologies.

Список литературы/References

1. Dewald O., Ren G., Duerr G.D., Zoerlein M., Klemm C., Gersch C., Tincey S., Michael L.H., Entman M.L., Frangogiannis N.G. Of mice and dogs: Species-specific differences in the inflammatory response following myocardial infarction. *Am J Pathol*. 2004;164:665-677
2. Voropaeva O.F., Tsgoev Ch.A., Yu.I. Shokin. Numerical simulation of the inflammatory phase of myocardial infarction. *Journal of Applied Mechanics and Technical Physics*. 2021;62(3):441-450
3. Voropaeva O.F., Tsgoev Ch.A. Numerical Modelling of Myocardial Infarction. I. Analysis of Spatiotemporal Aspects of the Local Inflammatory Response. *Math Biol Bioinf*. 2023;18(1):49-71
4. Voropaeva O.F., Tsgoev Ch.A. Numerical modelling of myocardial infarction. II. Analysis of macrophage polarization mechanism as a therapeutic target. *Math Biol Bioinf*. 2023;18(2):367-404
5. Voropaeva O.F., Bayadilov T.V. A mathematical model of aseptic inflammation dynamics. *J Appl Ind Math*. 2020;14:779-791
6. Mikhakhanova T.S., Voropaeva O.F. The trigger model of the dynamics of acute and chronic aseptic inflammation. *Math Biol Bioinf*. 2022;17(2):266-288
7. Yanenko N.N. *Fractional Step Method for Solving Multidimensional Problems of Mathematical Physics*. Novosibirsk. 1967

doi 10.18699/bgrs2024-12.2-02

Моделирование динамики острой фазы ВИЧ-инфекции

Гайнова И.

Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия
gajnova@math.nsc.ru

Ключевые слова: математическое моделирование; ВИЧ-инфекция; «ускользание» вируса; иммунный ответ

Мотивация и цель: Целью исследования является обзор работ по математическому моделированию острой фазы ВИЧ-инфекции и приложений результатов моделирования в клинической практике.
Методы и алгоритмы: Математические модели могут применяться для: исследования патогенетических механизмов ВИЧ-инфекции и реакции организма; проверки теоретических гипотез; прогнозирования течения ВИЧ-инфекции, кинетики ее развития (острая фаза, хроническое состояние, переход в стадию СПИДа); разработки персонализированной терапии; разработки и исследования новых стратегий лечения, составления сложных комбинаторных схем лечения; исследования механизмов «ускользания» вируса от лекарственных препаратов; обработки и анализа больших массивов данных – как клинических историй болезни, так и экспериментальных данных, полученных с помощью современных высокопроизводительных измерительных методов; оценки стоимости лечения и выработки рекомендаций на уровне правительств по предотвращению распространения болезни в странах с ограниченными ресурсами.
Результаты: В докладе критически рассмотрены математические модели острой фазы ВИЧ-инфекции, как базовые, так и современные, представленные в публикациях за последние 10 лет.
Выводы: Ряд исследований основывается на рассмотрении базовой (или стандартной) трехкомпонентной модели динамики острой фазы ВИЧ-инфекции, описывающей в терминах дифференциальных уравнений взаимодействие неинфицированных клеток-мишеней, инфицированных клеток и вирусных частиц. Такая модель позволяет представить качественную картину острой фазы: резкое увеличение вирусной нагрузки, ее пик и дальнейшее снижение до установления вирусного равновесия, вычислить базовую скорость репродукции вируса (R0). Величина R0 ($R0 < 1; > 1, << 1; >> 1$) является важным показателем течения ВИЧ-инфекции в дальнейшем. Для построения расширенных моделей динамики острой фазы в модель в качестве переменных включаются клетки-эффекторы, отвечающие за иммунный ответ (в большинстве работ рассматривается ответ CD8+ Т-клеток), различные типы клеток-мишеней, лекарственно-устойчивых и лекарственно-чувствительных штаммов ВИЧ, а также клеток-мишеней, инфицированных этими штаммами. Для идентификации неизвестных параметров модели применяются теория и методы решения обратных задач. В последние годы возрос интерес к математическому моделированию механизмов «ускользания» ВИЧ от иммунного ответа, таких как высокая генетическая изменчивость вируса вследствие ошибок обратной транскрипции, способность ВИЧ уже через 12 часов после инфицирования

подавлять экспрессию молекул MHC-I класса в инфицированных клетках и др. В ряде работ по моделированию острой фазы в исследование включается начальный период инфицирования – фаза эклипса (ранняя стадия внутриклеточной репликации вируса), что повышает эффективность модели.

Modeling of the acute phase HIV infection dynamics

Gainova I.
Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia
gajnova@math.nsc.ru

Key words: mathematical modeling; HIV infection; virus escape; immune responce

A number of studies are based on consideration of basic (or standard) three-component model of the dynamics of the acute phase of HIV infection, which describes in terms of differential equations, the interaction of uninfected target cells, infected cells and viral particles. This model allows us to present a qualitative picture of the acute phase: a sharp increase viral load, its peak and further decrease until the viral load is established equilibrium, calculate the basic rate of virus reproduction (R0). Magnitude R0 ($R0 < 1$; > 1, $<< 1$; $>> 1$) is an important flow indicator HIV infections in the future. To build advanced dynamics models acute phase, effector cells are included in the model as variables, responsible for the immune response (most studies consider the response CD8+ T cells), various types of target cells, drug-resistant and drug-sensitive strains of HIV, as well as target cells, infected with these strains. To identify unknown parameters of the model uses theory and methods for solving inverse problems. In recent years, interest in mathematical modeling of mechanisms has increased HIV “escape” from the immune response, such as high genetic variability of the virus due to reverse transcription errors, ability HIV suppresses the expression of molecules within 12 hours after infection MHC-I class in infected cells, etc. In a number of modeling works acute phase, the study includes the initial period of infection – eclipse phase (early stage of intracellular viral replication), which increases the efficiency of the model.

doi 10.18699/bgrs2024-12.2-03

Моделирование инфицирования в локальной атмосфере со случайной концентрацией вируса SARS-COV-2

Деревич И.*, Панова А.**

Московский государственный технический университет им. Н.Э. Баумана (национальный исследовательский университет), Москва, Россия

** DerevichIgor@bmstu.ru; ** Panova005@gmail.com*

Ключевые слова: клеточная модель; начальный иммунитет; поток вирионов; критическая концентрация патогена; эмпирическая функция плотности вероятности; время эвакуации

Мотивация и цель: Рост числа техногенных катастроф, террористических атак с возможным использованием биологически активных субстанций, например штаммов вирусных заболеваний, приводят к необходимости разработки методов численного моделирования заражения групп людей в экстремальных условиях. В работе в качестве прототипа вируса, оказывающего разрушающее действие на организм человека, выбран штамм SARS-COV-2 (Covid 19). Рассматривается следующий сценарий. В критических условиях случайным образом меняется относительное расстояние между индивидами, среди которых есть инфицированные. Также случайной является концентрация вирионов в локальной атмосфере. Цель работы – разработать метод прогноза эпидемиологического состояния индивидов после эвакуации из опасной зоны. В работе на основе модификации стандартной клеточной модели (см., например, [1]) рассмотрена динамика начальной стадии инфицирования в зараженной атмосфере.

Методы и алгоритмы: Стадия инфицирования (часы) существенно меньше времени развития тяжелой формы заболевания (дни). Динамика инфицирования определяется начальным уровнем иммунитета [2, 3]. На ранней стадии инфицирования в рамках стандартной клеточной модели предложено уравнение [4] для концентрации патогена, представленное ниже в безразмерном виде

$$\frac{\mathrm{d}X^*\left(t^*\right)}{\mathrm{d}t^*}=\Gamma^*X^*\left(t^*\right)\left[\frac{X^*\left(t^*\right)}{1+\alpha_{\mathrm{im}}^*X^*\left(t^*\right)}-1\right]-X^*\left(t^*\right)+J_{\mathrm{atm}}^*\left(t^*\right),\ X^*\left(0\right)=X_0^*.$$

Здесь X^* – концентрация патогена; J_{atm}^* – поток вирионов из атмосферы; Γ^* – скорость генерации клеток патогена из зараженных клеток организма; t^* – время; α_{im}^* – параметр, пропорциональный уровню начального иммунитета, $0<\alpha_{\mathrm{im}}<1$.

Уравнение баланса записано в предположении, что концентрация целевых клеток организма на ранней стадии практически не меняется. На основе анализа этого уравнения можно сделать следующие выводы, имеющие практическое значение. Существует критическая концентрация вирионов в организме, величина которой связана с начальным уровнем иммунитета:

$$X_{\mathrm{cr}}^*=\left[\left(1+\Gamma^*\right)/\Gamma^*\right]\left(1-\alpha_{\mathrm{im}}\right)^{-1}.$$

Если начальная концентрация патогена в организме выше критического значения $X_0^*>X_{\mathrm{cr}}^*$, то реализуется «полный» сценарий развития заболевания. Если

начальная концентрация вирионов меньше критического значения $X_0^* < X_{cr}^*$, то вирус вырождается естественным путем. Существует критическое значение потока вирионов из атмосферы. Если поток вирионов в организм меньше критического значения, то в организме устанавливается постоянная концентрация патогена, величина которой меньше критического значения. Для постоянного потока вирионов выше критического уровня реализуется монотонный рост концентрации патогена независимо от ее начального значения. В этом случае динамика заболевания после выхода из зараженной атмосферы определяется временем присутствия в опасной зоне. Если к моменту времени эвакуации концентрация патогена будет ниже критического уровня, то произойдет деградация вируса. Если концентрация патогена к моменту эвакуации из опасной зоны превысит критическое значение, то индивид будет в тяжелой форме заболевания.

Поток вирионов из зараженной атмосферы мы моделируем как логарифмически-нормальный случайный процесс. Прямое численное моделирование основано на системе стохастических обыкновенных дифференциальных уравнений (СОДУ), которая интегрируется современными методами типа Рунге–Кутта [5].

Результаты: Результаты прямого численного моделирования концентрации патогена в организме в течение времени пребывания в зараженной атмосфере и после эвакуации показаны на рис. 1 и 2.

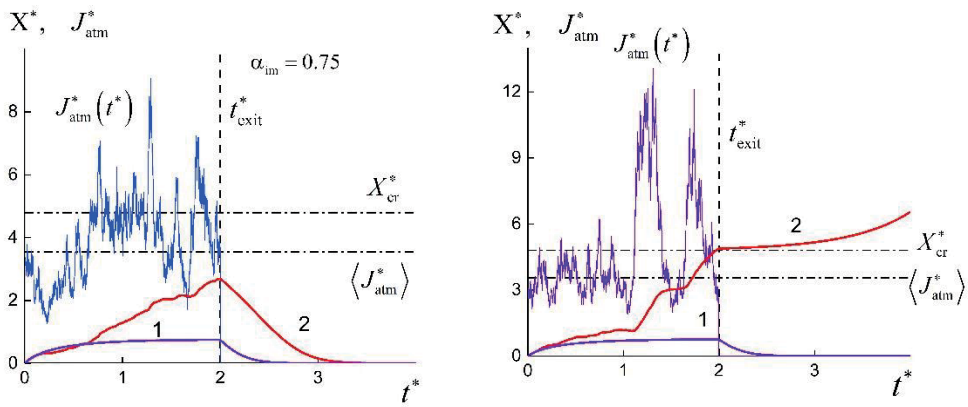


Рис. 1. Детерминированные (1) и случайные (2) изменения концентрации патогена в организме (t_{exit}^* – время эвакуации из зараженной зоны). Средний поток вирионов из атмосферы $\langle J_{atm}^* \rangle$ ниже критического уровня

Из рис. 1 и 2 можно заключить, что в атмосфере со случайной концентрацией вирионов могут реализоваться сценарии, не прогнозируемые детерминированной теорией. Например, при постоянном потоке вирионов из атмосферы ниже критического уровня индивиды, вышедшие из опасной зоны, должны восстановиться естественным путем. В случайной атмосфере с тем же осредненным значением потока реализуются сценарии, когда после эвакуации концентрация вирионов в организме будет выше критического значения (см. рис. 1). Если постоянный поток вирионов выше критического уровня и время пребывания индивидов в зараженной атмосфере достаточно для поглощения вирионов с

концентрацией выше критической, то в детерминированной теории после эвакуации вся группа должна перейти в тяжелую форму заболевания.

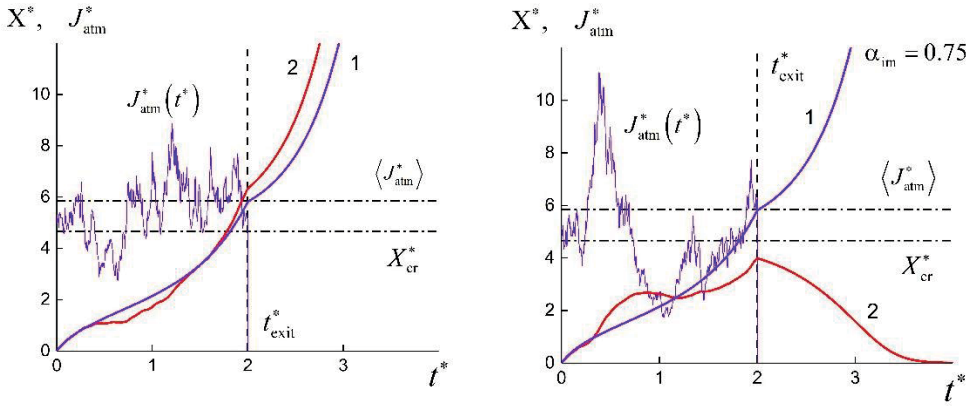


Рис. 2. Поток вирионов из атмосферы выше критического уровня. Остальные подписи как на рис. 1

В атмосфере со случайной концентрацией вирионов даже в этом случае будут сценарии, когда после эвакуации из опасной зоны концентрация будет ниже критического уровня и индивиды, вопреки ожиданиям, восстановятся (см. рис. 2). Оценку доли индивидов после эвакуации с тяжелой формой заболевания мы проводим на основе эмпирической функции плотности вероятности (ФПВ) случайной концентрации патогена в организме индивидов. Эмпирическая ФПВ рассчитывается на основе осреднения случайных сценариев для группы из 100 индивидов.

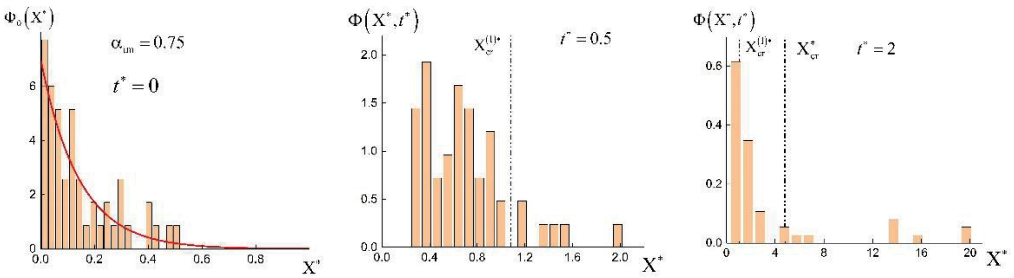


Рис. 3. Динамика изменения ФПВ распределения концентрации патогена для различных времен пребывания в зоне заражения. Средний поток вирионов в зараженной атмосфере ниже критического уровня

На рис. 3 и 4 представлена динамика ФПВ концентрации патогена в организме инфицированного при различных временах нахождения в зараженной зоне. Начальная ФПВ имеет узкое экспоненциальное распределение, сосредоточенное в области существенно ниже критической концентрации. На рис. 3 представлены ФПВ в различные моменты эвакуации. В детерминированном случае (см. рис. 1) при времени эвакуации $t^*_{\text{exit}} = 2$ все эвакуированные индивиды восстановятся. В атмосфере со случайной концентрацией вирионов будет заметная доля индивидов с тяжелой формой заболевания.

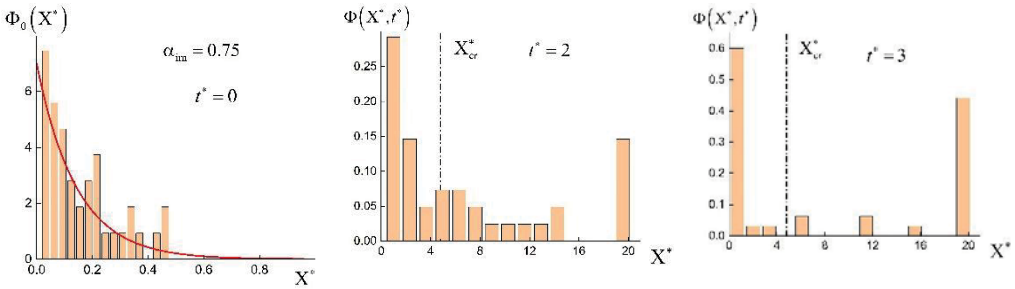


Рис. 4. Средний поток вирионов в зараженной атмосфере выше критического уровня. Остальные подписи как на рис. 3

Рисунок 4 иллюстрирует ФПВ концентрации патогена в различные моменты времени эвакуации для потока вирионов больше критического значения. В детерминированной теории (см. рис. 2) в момент времени $t^*_{\text{exit}} = 2$ вся эвакуированная группа будет в тяжелой форме заболевания. В атмосфере со случайной концентраций будет существенная доля индивидов, концентрация патогена в организме которых будет ниже критического значения. Даже при времени эвакуации $t^*_{\text{exit}} = 3$ остается заметная группа индивидов с низким уровнем концентрации патогена в организме.

Выводы: Представлен метод оценки доли инфицированных с тяжелой формой заболевания после эвакуации из зоны с зараженной атмосферой. В качестве модельного штамма выбран вирус SARS-COV-2. На основе численного решения системы СОДУ иллюстрируется качественное отличие начальной стадии инфицирования в детерминированном случае и в атмосфере со случайной концентрацией вирионов. Представлены результаты расчетов эмпирической ФПВ концентрации патогена в организме индивидов, эвакуированных из опасной зоны.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 23-29-00243 (https://rscf.ru/project/23-29-00243/).

Modeling of infection in a local atmosphere with a random concentration of SARS-COV-2 virus

Derevich I.*, Panova A.**

Bauman Moscow State Technical University (National Research University), Moscow, Russia

** DerevichIgor@bmstu.ru; ** Panova005@gmail.com*

Key words: cell model; initial immunity; virion flux; critical pathogen concentration; empirical probability density function; evacuation time

Motivation and Aim: The growing number of technogenic catastrophes, terrorist acts with possible application of biologically active substances, for example, strains of viral diseases lead to the need to develop methods of numerical modeling of infection of groups of people in extreme conditions. In this work, the SARS-COV-2 strain was chosen as a prototype virus with a devastating effect on the human body. The following scenario is considered. Under critical conditions, the relative distance between individuals, among which there are infected individuals, varies randomly. The

concentration of virions in the local atmosphere is also random. The aim of the work is to develop a method for predicting the epidemiological state of people after evacuation from a dangerous area. In this work, the dynamics of the initial stage of infection in a hazardous atmosphere was considered based on a modification of a standard cell model (see, e. g., [1]).

Methods and Algorithms: The stage of infection (hours) is much shorter than the development of severe disease (days). The dynamics of infection development is determined by the initial level of immunity [2, 3]. At an early stage, the standard cellular model is reduced to an equation [4] for pathogen concentration, represented in dimensionless form as follows

$$\frac{\mathrm{d}X^*\left(t^*\right)}{\mathrm{d}t^*} = \Gamma^* X^*\left(t^*\right)\left[\frac{X^*\left(t^*\right)}{1+\alpha_{\mathrm{im}}^* X^*\left(t^*\right)} - 1\right] - X^*\left(t^*\right) + J_{\mathrm{atm}}^*\left(t^*\right),\ X^*\left(0\right) = X_0^*.$$

Here X^* – concentration of the pathogen; J_{atm}^* – flux of virions from the atmosphere; Γ^* – rate of generation of pathogenic cells from infected cells of the organism; t^* – time; α_{im}^* – parameter proportional to the level of initial immunity $0 < \alpha_{\mathrm{im}} < 1$.

The balance equation is written under the assumption that the concentration of the target cells of the organism at the early stage is practically unchanged. Based on the analysis of this equation, the following conclusions of practical importance can be drawn There is a critical concentration of virions in the organism, the value of which is related to the initial level of immunity

$$X_{\mathrm{cr}}^* = \left[\left(1+\Gamma^*\right)/\Gamma^*\right]\left(1-\alpha_{\mathrm{im}}\right)^{-1}.$$

If the initial concentration of the pathogen in the organism is higher than the critical value $X_0^* > X_{\mathrm{cr}}^*$, the "complete" scenario of disease development is realized. If the initial concentration of virions is less than the critical value $X_0^* < X_{\mathrm{cr}}^*$, the virus naturally degenerates. There is a critical value of virion flux from the atmosphere. If the virion flux in the organism is less than the critical value, a constant concentration of the pathogen is established in the organism, which is less than the critical value. At a constant flux of virions above the critical level, a monotonic growth of pathogen concentration is realized regardless of its initial value. In this case, the dynamics of disease development after leaving the infected atmosphere is determined by the time of residence in the danger zone. If the concentration of pathogens is below the critical level by the time of evacuation, the virus will be degraded. If the concentration of pathogens exceeds the critical level by the time of evacuation from the danger zone, the person will be seriously ill. We model the flux of virions from an infected atmosphere as a lognormal random process. Direct numerical modeling is based on a system of stochastic ordinary differential equations (SODE), which is integrated by modern Runge–Kutta type methods [5].

Results: The results of direct numerical simulation of the pathogen concentration in the organism during the stay in the infected atmosphere and after evacuation are presented in Fig. 1 and Fig. 2. Figures 1 and 2 show that in an atmosphere with a random concentration of virions, scenarios not predicted by deterministic theory can be realized. For example, if there is a constant flux of virions from the atmosphere below a critical level, individuals that have left the danger zone should recover naturally. In a random atmosphere, scenarios are realized when the concentration of virions in the organism will be above a critical value after evacuation (Fig. 1). If there is a constant flux of virions

above the critical level, and the residence time in the infected atmosphere is sufficient to absorb virions with concentrations above the critical level, then according to deterministic theory the entire group should become seriously ill after evacuation. In an atmosphere with a random concentration of virions, even in this case, scenarios are possible, when after evacuation from the danger zone the concentration will be below the critical level and people, contrary to expectations, will recover (Fig. 2). We estimate the fraction of people after evacuation with severe disease based on an empirical probability density function (PDF) of the random pathogen concentration in an individual. The empirical PDF is calculated by averaging random scenarios for a group of 100 individuals.

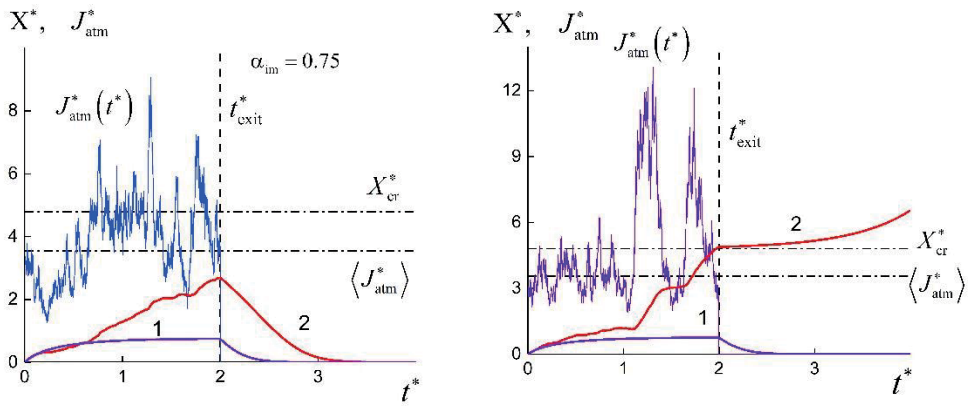


Fig. 1. Deterministic (1) and random pathogen concentrations (2) in the organism. Virion flux below the critical level

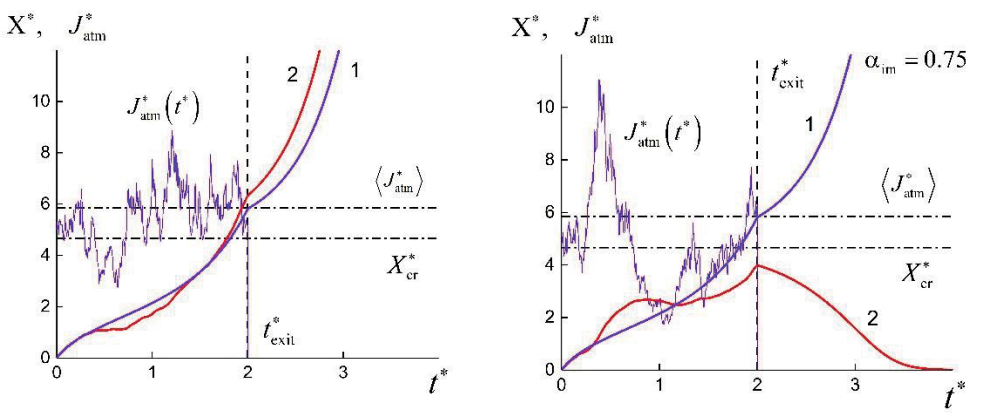


Fig. 2. Virion flux from the atmosphere above a critical level

Figures 3 and 4 show the dynamics of PDF of the pathogen concentration in the organism of an infected person at different residence times in the infected area. Fig. 3 shows the PDF at different evacuation times $t^*_{\text{exit}} = 2$. In the deterministic case (see Fig. 1), at the time of evacuation, all the evacuated persons recover. In an atmosphere with a random concentration of virions, there will be a noticeable proportion of people with severe illness. Figure 4 shows the PDF of pathogen concentration at various evacuation time points for virion fluxes exceeding the critical value. According to the deterministic theory (see Fig. 2), at some point in time $t^*_{\text{exit}} = 2$ the entire evacuation group will be

critically ill. In an atmosphere with a random concentration, there will be a significant fraction of people whose pathogen concentration will be below the critical value. Even at the time of evacuation $t^*_{\text{exit}} = 3$, there will still be a noticeable group of people with low pathogen concentrations.

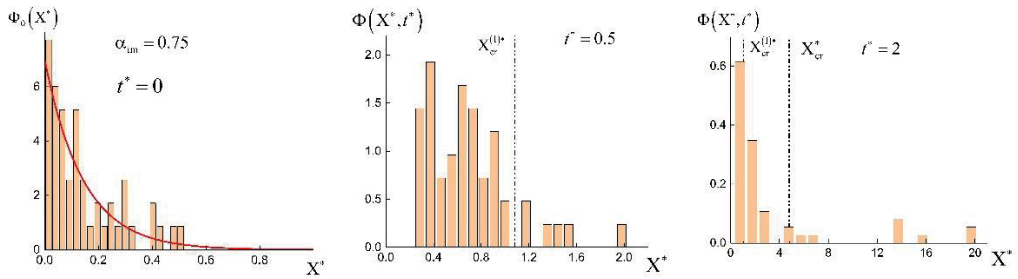


Fig. 3. Average virion flux below the critical level

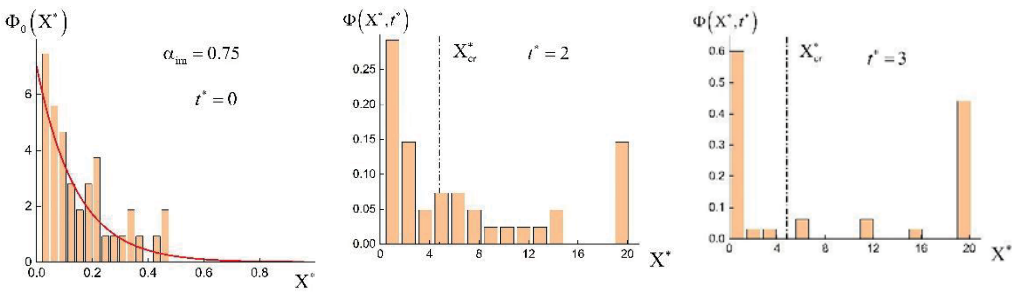


Fig. 4. Virion flux above the critical level

Conclusion: A method for estimating the fraction of infected people with severe disease after evacuation from an area with an infected atmosphere is presented. On the basis of the numerical solution of the SODE system, a qualitative difference between the initial stage of infection in the deterministic situation and in an atmosphere with a random concentration of virions is shown. The results of calculations of empirical PDF of the pathogen concentration in the organism of persons evacuated from the hazardous area are presented.

Funding: The study was funded by the Russian Science Foundation Grant No. 23-29-00243 (https://rscf.ru/project/23-29-00243/).

Список литературы/References

1. Kim K.S., Ejima K., Iwanami S. et al. A quantitative model used to compare within-host SARS-CoV-2, MERS-CoV, and SARS-CoV dynamics provides insights into the pathogenesis and treatment of SARS-CoV-2. *PLoS Biol*. 2021;19(3):e3001128. doi 10.1371/journal.pbio.3001128
2. Perelson A. Modelling viral and immune system dynamics. *Nat Rev Immunol*. 2002;2:28-36. doi 10.1038/nri700
3. Du S.Q., Yuan W. Mathematical modeling of interaction between innate and adaptive immune responses in COVID-19 and implications for viral pathogenesis. *J Med Virol*. 2020;92:1615-1628. doi 10.1002/jmv.25866
4. Деревич И.В., Панова А.А. Моделирование инфицирования в локальной атмосфере, зараженной вирусом SARS-COV-2. Стационарная концентрация вирионов. *Математическое моделирование*. 2024;36(3):67-86. doi 10.20948/mm-2024-03-05
5. Debrabant K., Rößler A. Classification of stochastic Runge–Kutta methods for the weak approximation of stochastic differential equations. *Math Comput Simul*. 2008;77(4):408-420

doi 10.18699/bgrs2024-12.2-04

Численное исследование начального этапа развития ВИЧ-1 инфекции в лимфатическом узле на основе стохастической модели и метода Монте–Карло

Логинов К.К.[1, 2]

[1] *Институт математики им. С.Л. Соболева СО РАН (Омский филиал), Новосибирск, Россия*

[2] *Институт вычислительной математики им. Г.И. Марчука РАН, Москва, Россия*

kloginov85@mail.ru

Ключевые слова: лимфатический узел; стохастическая модель; иммунология; ВИЧ-1 инфекция; метод Монте-Карло; вычислительный эксперимент

Мотивация и цель: Одним из активно развивающихся направлений математического моделирования в иммунологии является использование различных детерминированных и стохастических стадия-зависимых моделей. Детерминированные модели иммунных процессов обычно строятся на основе систем дифференциальных уравнений, включая уравнения с запаздыванием [1, 2]. Настоящая работа посвящена моделированию начального этапа динамики ВИЧ-1 инфекции в лимфатическом узле, в который проникло небольшое число вирусных частиц после инфицирования здорового человека в момент времени $t = 0$. Длительность промежутка моделирования $[0; T_{mod}]$ составляет несколько суток. В этом случае численность популяций вирусных частиц и инфицированных клеток достаточно мала, что делает не совсем пригодным применение известных детерминированных моделей для изучения динамики инфекции в указанный период. Для построения стохастической модели требуется использование целочисленных переменных, отражающих текущее количество вирусных частиц и клеток, а также дополнительных переменных, учитывающих предысторию формирования популяций вирусных частиц, инфицированных и продуктивно-инфицированных клеток [3, 4]. Целью работы является оценка вероятности завершения инфекционного процесса в зависимости от начального количества вирусных частиц и параметров модели, влияющих на распространение инфекции.
Методы и алгоритмы: При построении модели учитывается несколько факторов и событий, отражающих процесс развития ВИЧ-1 инфекции в лимфатическом узле [5]. Принято, что клетками-мишенями для вирусных частиц являются CD4+ Т-лимфоциты в состоянии покоя – клетки T_0. Клетки T_0 могут контактировать как с вирусными частицами V, так и с продуктивно-инфицированными клетками I_4 (клетками, производящими новые вирусные частицы V). В результате указанных контактов клетка T_0 превращается в зараженную клетку I_0. Клетка I_0 может покидать лимфатический узел вследствие миграционного оттока. Активация клеток I_0 к размножению происходит в результате их контактов с антиген-презентирующими клетками A. Численность популяций клеток T_0 и A в лимфатическом узле принята постоянной.
Активированная клетка I_0 превращается в клетку I_1, находящуюся в фазе G_1 клеточного цикла. После завершения фазы G_1 клеточного цикла клетка I_1

превращается соответственно в клетку I_2, I_3, I_+ с вероятностями α_{I_2}, α_{I_3}, α_{I_+}, $\alpha_{I_2} + \alpha_{I_3} + \alpha_{I_+} = 1$. Здесь:

- I_2 – зараженные клетки-мишени, находящиеся в фазах $S-G_2-M$ клеточного цикла;
- I_3 – зараженные клетки-мишени, остановленные в фазе G_2 клеточного цикла;
- I_+ – зараженные клетки-мишени, находящиеся в фазах $S-G_2-M$ клеточного цикла и продолжающие размножение за счет деления, но они и их потомки не способны к производству вирусных частиц (эти клетки в модели не рассматриваются).

Клетка I_2 спустя время $\omega_{I_2} = const > 0$ после своего появления завершает деление и образует две продуктивно-инфицированные клетки I_4. Клетка I_3 спустя время $\omega_{I_3} = const > 0$ после своего появления превращается в продуктивно-инфицированную клетку I_4 (без деления). Выражение $2\,\alpha_{I_2} + \alpha_{I_3}$ означает среднюю численность продуктивно-инфицированного потомства I_4 от одной клетки I_1.

Клетка I_4 не способна к делению. Клетка I_4 после своего появления в течение периода длительностью ω_{I_4} производит новые вирусные частицы V с интенсивностью $\eta > 0$. Величина ω_{I_4} является случайной, ее распределение задается функцией дожития

$$P(\omega_{I_4} \geq a) = L_{I_4}(a), \;\; 0 \leq a < \infty, \qquad L_{I_4}(0) = 1.$$

При завершении периода длительностью ω_{I_4} клетка I_4 теряет свою продуктивную способность и в дальнейшем погибает под влиянием разрушительных для нее процессов, обусловленных производством вирусных частиц.

Вирусная частица V может покидать лимфатический узел вследствие миграционного оттока, а также может погибнуть в результате воздействия некоторых факторов. В модели не учитывается уменьшение численности популяций зараженных вирусами клеток I_0, I_1, I_2, I_3 вследствие естественного старения и гибели под воздействием вирусного заражения. На рис. 1 представлена схема модели.

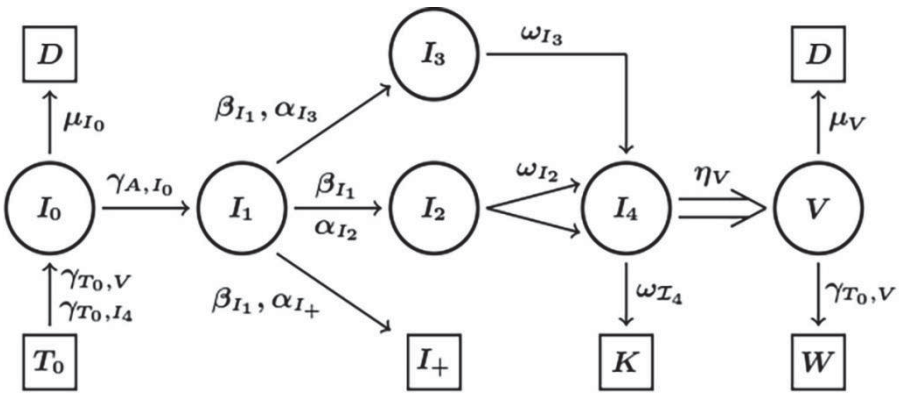


Рис. 1. Схема модели

В работе приведено описание модели в форме случайного процесса, переменными которого в каждый фиксированный момент времени являются: 1) численности популяций клеток и вирусных частиц – неотрицательные целочисленные случайные величины; 2) семейства уникальных типов клеток, учитывающие предысторию формирования популяций клеток I_2, I_3, I_4. Представлен алгоритм численного моделирования на основе метода Монте-Карло [6, 7].
Результаты: Исследована динамика популяций вирусных частиц и клеток, получены оценки вероятности искоренения ВИЧ-1 инфекции в лимфатическом узле для различных значений коэффициента распространения инфекции R_0, начальной численности вирусных частиц V_0, а также различных законов распределения случайной величины ω_{I_4} с одним и тем же математическим ожиданием $E\omega_{I_4}$. Параметры модели выбирались в соответствии с работой [4], интервальные оценки численностей клеток, вирусных частиц и вероятности искоренения инфекции получены по выборке из 10 000 реализаций случайного процесса на уровне доверия 0.99. Для задания функции дожития $L_{I_4}(a)$ использовался экспоненциальный закон распределения, а также рассматривались степенное и равномерное на некотором конечном интервале распределения величины ω_{I_4}.
Выводы: Вычисления показывают, что для используемых распределений оценки вероятности искоренения инфекции в лимфатическом узле различаются в основном для экспоненциального и остальных распределений. Кроме того, при экспоненциальном законе распределения случайной величины ω_{I_4} основная часть траекторий общей численности инфицированных клеток и вирусных частиц попадает в нулевое поглощающее состояние за более короткий промежуток времени, чем при равномерном и степенном законах распределения.
Финансирование: Исследование поддержано Российским научным фондом (проект № 23-11-00116).

Numerical study of the initial stage of spread of HIV-1 infection in a lymphatic node based on a stochastic model and the Monte Carlo method

Loginov K.K.[1, 2]

[1] *Sobolev Institute of Mathematics (Omsk Branch), SB RAS, Novosibirsk, Russia*

[2] *Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia*

kloginov85@mail.ru

Key words: lymphatic node; stochastic model; immunology; HIV-1 infection; Monte Carlo method; computational experiment

Motivation and Aim: One of the actively developing directions of mathematical modeling in immunology is the use of various deterministic and stochastic stage-dependent models. Deterministic models of immune processes are usually research on the base of differential equations systems, including equations with delays [1, 2]. This work is devoted to modeling the initial period of the dynamics of HIV-1 infection in a lymphatic node into which a few numbers of viral particles have penetrated after

infection of a healthy individual at time $t = 0$. Simulation interval duration $[0;\ T_{mod}]$ is several days. In this case, the population size of viral particles and infected cells is quite small, and known deterministic models not entirely suitable for studying the dynamics of infection during the concerned period. To build a stochastic model, it is necessary to use integer variables reflecting the current quantity of viral particles and cells, as well as additional variables that take into account the history of the formation of populations of viral particles, infected and productively infected cells [3, 4]. The aim of the work is to estimate the probability of infectious process termination depending on the initial number of viral particles and model parameters influencing the spread of infection.

Methods and Algorithms: Several factors and events taking into account the development of HIV-1 infection in the lymphatic node is reflected for construction of model [5]. We assume that the target cells for viral particles are CD4+ T lymphocytes in a resting state – T_0 cells. Cells T_0 can contact both virus particles V and productively infected cells I_4 (cells producing new virus particles V). As a result of these contacts, T_0 cell become the infected I_0 cell. Cell I_0 may leave the lymphatic node due to migratory outflow. Activation of cells I_0 to reproduce occurs as a result of their contacts with antigen-presenting cells A. The population size of T_0 and A cells in the lymphatic node is assumed to be constant.

The activated cell I_0 become the cell I_1 in the G_1 phase of the cell cycle. After the completion of phase G_1 of the cell cycle, the cell I_1 become the cell I_2, I_3, I_+, respectively, with probabilities α_{I_2}, α_{I_3}, α_{I_+}, $\alpha_{I_2} + \alpha_{I_3} + \alpha_{I_+} = 1$. Designations:

- I_2 – infected target cells in the $S-G_2-M$ phases of the cell cycle;
- I_3 – infected target cells aborted in the G_2 phase of the cell cycle;
- I_+ – infected target cells in the $S-G_2-M$ phases of the cell cycle and continue to reproduce through fission, but they and their descendants are not capable of producing viral particles (these cells are not considered in the model).

The cell I_2 after time $\omega_{I_2} = const > 0$ after its appearance, completes division and forms two productively infected cells I_4. The cell I_3 after time $\omega_{I_3} = const > 0$ after its appearance, become a productively infected cell I_4 (without division). Expression $2\,\alpha_{I_2} + \alpha_{I_3}$ means the average number of productively infected progeny I_4 from one cell I_1.

The cell I_4 is not capable of division. The cell I_4 after its appearance during a period of duration ω_{I_4} produces new viral particles V with intensity $\eta > 0$. The variable ω_{I_4} is random and its distribution is specified by the survival function

$$P(\omega_{I_4} \geq a) = L_{I_4}(a),\ \ 0 \leq a < \infty, \qquad L_{I_4}(0) = 1.$$

At the end of a period of duration ω_{I_4} the cell I_4 loses its productive ability and subsequently dies under the influence of destructive processes as the result of the production of viral particles.

Viral particle V can leave the lymphatic node due to migratory outflow and can also die as a result of exposure to some factors. The model does not take into account the decrease in the population of virus-infected cells I_0, I_1, I_2, I_3 owing to natural aging and death under the influence of viral infection.

The diagram of the model is shown in the Fig. 1.

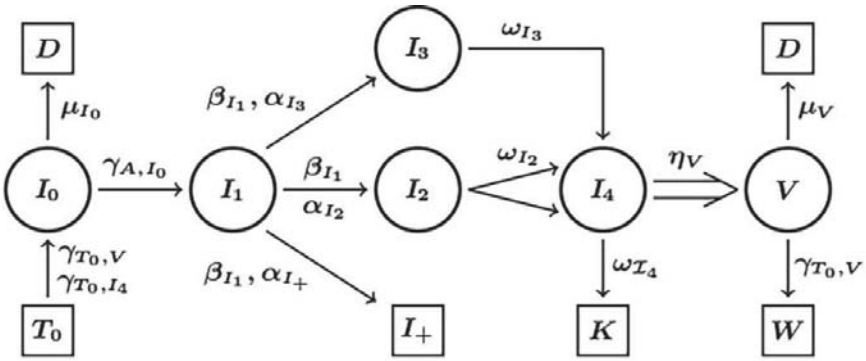


Fig. 1. Scheme of the model

Description of the model in the form of a random process is presented at work. The variables of this process at each fixed time point are: (1) the number of populations of cells and viral particles – non-negative integer random variables; (2) families of unique cell types taking into account the history of the formation of cell populations I_2, I_3, I_4. A numerical simulation algorithm based on the Monte Carlo method is presented [6, 7].
Results: The dynamics of populations of viral particles and cells is studied, and estimates of the probability of HIV-1 infection eradication in the lymphatic node is obtained for various values of the infection spread coefficient R_0, the initial quantity of viral particles V_0 as well as various laws of distribution of the random variable ω_{I_4} with the same mathematical expectation $E\omega_{I_4}$. The model parameters are chosen in accordance with [4], the numbers of cells, viral particles and the probabilities of infection eradication are estimated from a sample of 10000 realizations of the random process at a confidence level of 0.99. To set the survival function $L_{I_4}(a)$, the exponential distribution law is used, and also power law and uniform distribution on a some finite interval of the variable ω_{I_4} is considered.
Conclusion: Calculations show that for the specified distributions, estimates of the probability of infection eradication in a lymphatic node differ mostly for the exponential and other distributions. In addition, for the exponential law of distribution of the random variable ω_{I_4}, the main part of the trajectories of the total number of infected cells and viral particles get to zero absorbing state in a shorter period of time than for the uniform and power law of distribution.
Funding: The study is supported Russian Science Foundation (project No. 23-11-00116).

Список литературы/References

1. Bocharov G. et al. Human Immunodeficiency Virus Infection: From Biological Observations to Mechanistic Mathematical Modelling. *Math Model Nat Phenom.* 2012;7(5):78-104. doi 10.1051/mmnp/20127507
2. Черешнев В.А. и др. Введение в задачи моделирования и управления динамикой ВИЧ инфекции. М.; Ижевск: Институт компьютерных исследований, 2016
[Chereshnev V.A. et al. Introduction to the problems of modeling and managing the dynamics of HIV infection. Moscow; Izhevsk: Institute of Computer Research, 2016 (in Russian)]
3. Bocharov G.A., Loginov K.K., Pertsev N.V., Topchii V.A. Direct statistical modeling of HIV-1 infection based on a non-Markovian stochastic model. *Comp Math Math Phys.* 2021;61(8):1229-1251. doi 10.1134/S0965542521060026
4. Pertsev N.V., Loginov K.K. Stochastic Modeling in Immunology Based On a Stage-Dependent Framework with Non-Markov Constraints for Individual Cell and Pathogen Dynamics. *Math Biol Bioinf.* 2023;18(2):543-567. doi 10.17537/2023.18.543

5. Аббас А.К., Лихтман Э.Г, Пиллаи Ш. Основы иммунологии. Функции иммунной системы и их нарушения. М.: Гэотар-Медиа, 2022
[Abbas A.K., Likhtman E.G., Pillai Sh. Fundamentals of immunology. Functions of the immune system and their disorders. Moscow: Geotar-Media, 2022 (in Russian)]
6. Михайлов Г.А., Войтишек А.В. Численное статистическое моделирование. Методы Монте-Карло. М.: Академия, 2006
[Mikhailov G.A., Voitishek A.V. Numerical Statistical Simulation. Monte-Carlo Methods. Moscow: Akademia, 2006 (in Russian)]
7. Marchenko M. PARMONC – A Software Library for Massively Parallel Stochastic Simulation. In: Malyshkin V. (Ed.). Parallel Computing Technologies. Lecture Notes in Computer Science. Vol. 6873. Springer, 2011;302-316. doi 10.1007/978-3-642-23178-0_27

doi 10.18699/bgrs2024-12.2-05

Алгоритмы реализации индивидуальной клеточной динамики в плотно упакованных средах

Савинков Р.[1, 2]

[1] *Первый МГМУ имени И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия*

[2] *Институт вычислительной математики имени Г.И. Марчука Российской академии наук (ИВМ РАН), Москва, Россия*

r.savinkov@inm.ras.ru

Ключевые слова: иммунология; клетка; моделирование; стохастические процессы

Мотивация и цель: Множество иммунных процессов, происходящих в организме, связано с взаимодействием малого количества агентов (вирусных частиц, клеток иммунной системы). В частности, на начальной стадии процесс заражения вирусом ВИЧ-1 в лимфатических узлах человека задействует считанные единицы или десятки клеток в условиях плотно упакованной различными объектами (клетками, коллагеновыми структурами) Т-клеточной зоны. В связи с потребностью исследования случайных процессов в таких расчётных областях, а также схожих ситуациях, были разработаны два подхода к моделированию клеток и их динамических свойств в виде структур вокселей, связанных между собой и алгоритмов, управляющих изменением этих структур.

Методы и алгоритмы: Для реализации поставленной задачи были применены два подхода:

1) метод паттернов: деформации одобрялись или отвергались на основании предвычисленного списка шаблонов, контролирующих выполнение двух условий: клетка вследствие деформации не может потерять структурную целостность, а также не может поглотить иные объекты внутри моделируемой области.

2) метод поиска на графе: после оценки наиболее предпочтительной точки деформации тела клетки, все дальнейшие решения по совершаемым с вокселями манипуляциям принимались на основании дальности рассматриваемых элементов от выбранной в качестве предпочтительной точки.

Результаты: В результате проделанной работы было показано, что предложенные алгоритмы позволяют частично воспроизвести статистические свойства индивидуальной клеточной динамики наблюдаемых *in vitro* клеток иммунной системы и позволяют выполнять моделирование поведения порядка 100000 клеток в плотно упакованной (более 80 %) среде.

Выводы: Разработанные алгоритмы могут быть применены для моделирования движения отдельных клеток в плотно упакованных средах в различных задачах математического моделирования, а именно, в ситуациях, когда количество значимых для процесса агентов существенно меньше общего числа объектов, присутствующих в расчётной области. Оба алгоритма обладают как преимуществами, так и недостатками, так что выбор в пользу одного из них (или выбор иного подхода к решению) существенно зависит от цели исследования.

Финансирование: Исследование поддержано грантом РНФ 23-11-00116.

Algorithms for implementing individual cell dynamics in dense packed environments

Savinkov R.[1, 2]

[1] *Sechenov First Moscow State Medical University, Ministry of Health of the Russian Federation (Sechenov University), Moscow, Russia*

[2] *Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences (INM RAS), Moscow, Russia*
r.savinkov@inm.ras.ru

Key words: immunology; cell; modeling; stochastic processes

Motivation and Aim: Many immune processes occurring in the body involve interactions of a small number of agents (viral particles, immune cells). For instance, in the initial stage of HIV-1 virus infection process in human lymph nodes, only a few or tens of cells are involved in densely packed conditions with various objects (cells, collagen structures) in the T-cell zone. To study random processes in such computational domains and similar situations, two approaches have been developed for modeling cells and their dynamic properties as voxel structures interconnected and algorithms governing the changes in these structures.
Methods and Algorithms: Two approaches were applied to achieve the set goal:
1) Pattern method: deformations were accepted or rejected based on a precomputed list of patterns controlling the fulfillment of two conditions: the cell cannot lose structural integrity due to deformation and cannot engulf other objects within the modeled area.
2) Graph search method: after evaluating the most preferred point of cell body deformation, further decisions on voxel manipulations were made based on the distance of considered elements from the selected preferred point.
Results: The study showed that the proposed algorithms partially reproduce the statistical properties of individual cell dynamics observed *in vitro* in immune system cells and enable modeling the behavior of around 100,000 cells in a densely packed (over 80 %) environment.
Conclusion: The developed algorithms can be applied to model the movement of individual cells in densely packed environments in various mathematical modeling tasks, specifically in situations where the number of agents significant to the process is substantially less than the total number of objects present in the computational domain. Both algorithms have their own advantages and disadvantages, so the choice between them (or choosing a different approach to solving the problem) significantly depends on the research goal.
Funding: The research was supported by the Russian Science Foundation grant 23-11-00116.

doi 10.18699/bgrs2024-12.2-06

Вычисление и анализ стационарных решений модели динамики инфекции COVID-19

Христиченко М.[1]*, Нечепуренко Ю.[1], Гребенников Д.[1, 2], Бочаров Г.[1, 2]
[1] *Институт вычислительной математики им. Г.И. Марчука РАН, Москва, Россия*
[2] *Первый Московский государственный медицинский университет им. И.М. Сеченова Министерства здравоохранения Российской Федерации, Москва, Россия*
* *misha.hrist@gmail.com*

Ключевые слова: дифференциальные уравнения с запаздыванием; стационарные решения; методы вычислительной алгебры; анализ устойчивости; трассировка по параметру

Мотивация и цель: Одним из широко применяемых в настоящее время теоретических подходов к исследованию возможных хронических форм инфекционных заболеваний и разработке методов их лечения является использование математических моделей динамики заболеваний и иммунного ответа, представляющих собой системы дифференциальных уравнений с запаздыванием. Устойчивые нетривиальные стационарные решения таких моделей с достаточно высокой вирусной нагрузкой можно интерпретировать как хронические формы заболевания. Поэтому для исследования таких форм заболеваний необходимо вычислять и выполнять анализ стационарных решений систем с запаздыванием. В работах [1–4] авторами доклада были предложены методы вычисления стационарных решений систем с запаздыванием, исследования их устойчивости и их трассирования по параметрам моделей (исследования зависимости от параметров моделей). Ранее эти методы были успешно использованы для вычисления и анализа стационарных решений двух широко известных математических моделей вирусных инфекций: модели динамики экспериментальной инфекции, вызванной вирусами лимфоцитарного хориоменингита [5], и модели инфекции человека вирусами гепатита В, являющейся калиброванной версией модели противовирусного иммунного ответа Марчука–Петрова [6]. В частности, были впервые вычислены стационарные решения этих моделей, соответствующие хроническим формам заболеваний. Также впервые были найдены значения параметров моделей, при которых у них есть свойства гистерезиса (зависимости состояния модели от предыстории при изменении значений параметров) и бистабильности (наличие при одних и тех же значениях параметров двух устойчивых стационарных решений). Данный доклад посвящен применению разработанных методов для анализа стационарных решений модели динамики инфекции COVID-19, предложенной в [7].

Методы и алгоритмы: В работах [1–4] авторами доклада были предложены методы вычисления стационарных решений систем с запаздыванием, исследования их устойчивости и их трассирования по параметрам систем. Для вычисления стационарных решений предлагается использовать методы вычислительной алгебры, в частности реализованные в процедуре NSolve пакета Mathematica. Исследование асимптотической устойчивости стационарного решения сводится к решению нелинейной проблемы собственных значений с линейным и экспоненциальным вхождением спектрального параметра и к

проверке того, что все найденные собственные значения лежат строго в левой полуплоскости. Ведущие собственные значения вычисляются методом последовательных линейных проблем. Для вычисления начальных приближений исходная нелинейная проблема собственных значений аппроксимируется рациональной проблемой собственных значений, которая, в свою очередь, сводится к полиномиальной. Одновременное трассирование всех вещественных стационарных решений по параметру сводится к построению подмножеств некоторой действительной алгебраической кривой, на которых стационарные решения являются однозначными функциями варьируемого параметра.
Результаты: С помощью предложенных методов были вычислены стационарные решения модели COVID-19, соответствующие хроническим формам заболевания различной тяжести. Выполнен анализ зависимости стационарных решений модели и их устойчивости от значений различных параметров модели.
Выводы: Предложенные методы позволяют эффективно вычислять стационарные решения модели COVID-19 и выполнять их анализ. Эти методы позволяют эффективно искать в областях значений параметров модели те значения, при которых возможно хроническое течение заболевания, а также значения, при которых у модели появляются такие важные свойства как бистабильность и гистерезис. Полученные результаты позволяют глубже понять механизмы патогенеза неблагоприятных форм течения инфекции, вызванной вирусами SARS-CoV-2, и необходимы для дальнейшего анализа возможности построения режимов комбинированных терапий инфекции COVID-19.
Финансирование: Работа выполнена при финансовой поддержке Российского научного фонда, проект № 22-11-00025.

Computation and analysis of stationary solutions of the COVID-19 infection dynamics model

Khristichenko M.[1]*, Nechepurenko Yu.[1], Grebennikov D.[1, 2], Bocharov G.[1, 2]

[1] *Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia*

[2] *Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation, Moscow, Russia*

* *misha.hrist@gmail.com*

Key words: delay differential equations; stationary solutions; computer algebra methods; stability analysis; tracing along the parameter

Motivation and Aim: Currently, one of the most common theoretical approaches to studying potentially possible chronic forms of infectious diseases and developing ways to treat them is the use of mathematical models of disease and immune response dynamics represented by systems of delay differential equations. Stable nontrivial stationary solutions of such models with sufficiently high viral load can be interpreted as chronic forms of disease. Therefore, to study such forms of diseases, it is essential to compute and perform analysis of stationary solutions of time-delay systems. In [1–4], the authors of the report proposed methods for computing stationary solutions of time-delay systems, analyzing their stability, and tracing them along the model parameters (analyzing dependence from the model parameters). Earlier, these methods were successfully used to compute and analyze stationary solutions of two widely known

mathematical models of viral infections: the model of the dynamics of experimental infection caused by lymphocytic choriomeningitis viruses [5] and the model of human infection by hepatitis B viruses, which is a calibrated version of the Marchuk–Petrov antiviral immune response model [6]. In particular, the stationary solutions of these models corresponding to chronic forms of diseases were computed for the first time. Moreover, for the first time the values of the model parameters at which the models exhibit the properties of hysteresis (dependence of the model state on the prehistory while changing the parameter values) and bistability (presence of two stable stationary solutions at the same parameter values) were found. This report is devoted to the application of the developed methods for analyzing stationary solutions of the COVID-19 infection dynamics model proposed in [7].

Methods and Algorithms: In [1–4], the authors of the report proposed methods for computing stationary solutions of time-delay systems, analyzing their stability, and tracing them by system parameters. To compute stationary solutions, it is proposed to use methods of computer algebra, in particular those implemented in the NSolve procedure of Mathematica. The study of asymptotic stability of the stationary solution is reduced to solving the nonlinear eigenvalue problem with linear and exponential entry of the spectral parameter, and checking that all found eigenvalues lie strictly in the left half-plane. The leading eigenvalues are computed by the method of successive linear problems. To compute the initial approximations, the original nonlinear eigenvalue problem is approximated by a rational eigenvalue problem, which in turn reduces to a polynomial one. Simultaneous tracing of all real stationary solutions along the parameter reduces to constructing subsets of some real algebraic curve on which the stationary solutions are single-valued functions of the varied parameter.

Results: Stationary solutions of the COVID-19 model corresponding to chronic forms of the disease of different severity were computed using the proposed methods. The dependence of stationary solutions of the model and their stability from the values of various model parameters was analyzed.

Conclusion: The proposed methods allow one to efficiently compute stationary solutions of the COVID-19 model and analyze them. These methods allow one to efficiently search in the regions of the model parameter values for values at which the chronic course of the disease is potentially possible, as well as for values at which the model can exhibit such important properties as bistability and hysteresis. The obtained results allow for deeper understanding of mechanisms behind the pathogenesis of adverse of the course of infection caused by SARS-CoV-2 viruses and are essential for further research into potential for developing combined therapy regimens for COVID-19 infection.

Funding: This work was supported by the Russian Science Foundation (Grant 22-11-00025).

Список литературы/References

1. Nechepurenko Yu.M., Khristichenko M.Yu., Grebennikov D.S., Bocharov G.A. Bistability analysis of virus infection models with time delays. *Discrete Contin. Dyn. Syst. Ser. S.* 2020;13(9):2385-2401
2. Khristichenko M.Yu., Nechepurenko Yu.M., Grebennikov D.S., Bocharov G.A. Modelling chronic hepatitis B using the Marchuk-Petrov model. *J Phys Conf Ser*. 2021;2099:012036
3. Христиченко М.Ю., Нечепуренко Ю.М., Гребенников Д.С., Бочаров. Г.А. Численный анализ стационарных решений систем с запаздывающим аргументом в математической иммунологии. *Современная математика. Фундаментальные направления*. 2022;68(4):686-703
4. Sklyarova E.V., Nechepurenko Yu.M., Bocharov G.A. Numerical steady state analysis of the Marchuk–Petrov model of antiviral immune response. *Russ J Numer Anal Math Modell*. 2020;35(2):95-110

5. Bocharov G.A. Modelling the dynamics of LCMV infection in mice: conventional and exhaustive CTL responses. *J Theor Biol.* 1998;192(3):283-308
6. Marchuk G.I., Petrov R.V., Romanyukha A.A., Bocharov G.A. Mathematical model of antiviral immune response. I. Data analysis, generalized picture construction and parameters evaluation for hepatitis B. *J Theor Biol.* 1991;151(1):1-40
7. Grebennikov D., Karsonova A., Loguinova M., Casella V., Meyerhans A., Bocharov G. Predicting the Kinetic Coordination of Immune Response Dynamics in SARS-CoV-2 Infection: Implications for Disease Pathogenesis. *Mathematics*. 2022;10:3154

doi 10.18699/bgrs2024-12.2-07

Frontier modelling challenges in mathematical immunology

Bocharov G.[1, 2]

[1] *Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia*

[2] *Institute of Computer Science and Mathematical Modelling, Sechenov University, Moscow, Russia*

g.bocharov@inm.ras.ru

Key words: mathematical modelling; immune system; cellular ensembles; regulatory networks; intracellular regulation; evolutionary dynamics; information complexity; infectious

Motivation and Aim: In description, analysis and prediction of infectious diseases that depend on the reactions of the immune system a range of challenges exist. These include the curse of dimensionality of the immune system state (clonal repertoire) space, the multiplicity of dynamics trajectories of pathological processes, the nonlinearity of regulation loops, and the heterogeneity and variability of innate- and adaptive immunity [1]. To overcome these challenges, the deployment of novel mathematical models and computational approaches is needed. These analytical tools should enable a multi-physics description of the immune system dynamics in the high-dimensional space of physical and phenotypic traits, allow a cause-and-effect type of analysis of infectious disease pathogenesis and a robust quantitative predictions of the host immune system reaction to combination therapies.

Methods and Algorithms: To build mathematical models that meet the requirements of the current level of research in immunology, the distributed parameters systems in the space of phenotypic traits (e.g., affinity/avidity of receptors) and physical characteristics, the network models of interaction of cellular ensembles, structural modules of intracellular regulation are developed [1]. New elements of the presented multi-physics modelling approach and the developed analysis tools are as follows: (1) the use of methods of evolutionary dynamics on adaptive landscapes to describe the connectivity of populations of immune cells and the clonal repertoire under the influence of random antigenic forcing, (2) the formation of fitness landscapes to assess the information-entropy characteristics of the system and predict changes in its complexity and system efficiency, (3) the description of a hierarchical organization of regulatory processes, and (4) the application of meta-analysis algorithms for calibration of the processes described in models. The numerical implementation of the models requires elaboration of problem-oriented methods and algorithms. In particular, a geometric model of the murine lymphatic system has been developed and characterized in terms of its structural organization (graph topological properties) and the lymph transfer function [2].

Results: A mathematical model of lymphocytic choriomeningitis virus (LCMV) replication in infected cells has been developed and calibrated [3]. The murine LCMV infection is a fundamental model system in experimental immunology. By studying the sensitivity of the mathematical model, potential targets for the development of antiviral drugs have been identified. Processes that determine the development of experimental LCMV infection into the acute or chronic phenotypes were examined. Mathematical model describing the turnover kinetics of exhausted CD8 T cell population in chronic LCMV infection has been constructed and used to make predictions of the effect of anti-PDL-1 therapies. Based on the systems biology concept of multi-stability and the

prediction of multiplicative cooperativity between virus-specific cytotoxic T cells and neutralising antibodies, we argue for the requirements to engage multiple immune system components for functional cure strategies. The arguments are derived from LCMV model system studies and are translated to HIV-1 infection [4].

Conclusion: The fundamental result of our study is the elaboration of new approaches to modelling the functioning of the immune system of humans and experimental animals and methods for analyzing its "complexity". New classes of mathematical and computer models are being developed enabling a mechanistic description of the network and repertoire structures, hierarchical regulation, and evolutionary dynamics of immune processes under normal conditions and in infectious diseases.

Funding: The study is funded by the Russian Science Foundation (grant No. 23-11-00116, https://rscf.ru/project/23-11-00116/).

References

1. Bocharov G. et al. Multiphysics modelling of immune processes using distributed parameter systems. *Russ J Numer Anal Math Modelling*. 2023;38(5):279-292
2. Grebennikov D. et al. Network Modeling of Murine Lymphatic System. *Algorithms*. 2023;16(3):168
3. Sergeeva J. et al. Mathematical Model Predicting the Kinetics of Intracellular LCMV Replication. *Mathematics*. 2023;11(21):4454
4. Bocharov G. et al. Functional cure of a chronic virus infection by shifting the virus – host equilibrium state. *Front Immunol*. 2022;13:904342

doi 10.18699/bgrs2024-12.2-08

Computational methods for hybrid stochastic multiscale modelling of virus infection dynamics

Grebennikov D.[1, 2]

[1] *Marchuk Institute of Numerical Mathematics, RAS, Moscow, Russia*

[2] *World-Class Research Center 'Digital Biodesign and Personalized Healthcare', Sechenov First Moscow State Medical University, Moscow, Russia*

grebennikov_d_s@staff.sechenov.ru

Key words: multiscale modelling; hybrid agent-based model; stochastic model; Markov chains with time-inhomogeneous propensities; Monte Carlo simulation algorithms; intracellular virus replication; immune cell motility; virus infection dynamics

Motivation and Aim: The dynamics of virus infections are governed by a variety of processes at multiple scales: at the whole organism level (systemic transport of immune cells, cytokines and viruses between lymphoid and target organs via blood and lymphatic systems), at the tissue level (immune cell locomotion, virus and cytokine spread in lymphoid and virus sensitive tissues, cell-to-cell virus transmission), and at the intracellular level (intracellular virus replication and molecular signalling determining the immune cell states). Certain processes at several scales are stochastic, and the resulting randomness influences other scales. By adopting an integrative approach, we have developed computational methods for studying virus infections through hybrid stochastic multi-scale multi-compartment modelling [1, 2].

Methods and Algorithms: Transport of viruses, cells and molecules between infection sites, lymphoid tissues and blood is modelled using a multi-compartment approach. Reaction-diffusion equations are used to model diffusion of free virions in the lymphoid tissue with moving sources, which are determined by the positions of the infected cells (immune cell motility model) and the rate of virion secretion from these cells (intracellular model). These transport equations are solved numerically using the finite element method on the unstructured tetrahedral meshes over the generalized lymph node geometry adaptively refined to the spherical sub-domains occupied by the productively infected cells [1]. Immune and target cells are treated as individual agents, with their positions determined by the cell motility model that parametrises the intercellular interaction forces, friction and the stochastic forces of active cellular motility [3]. The equations of cell motion are integrated numerically using the symplectic Euler method. Intracellular models describe the production of the virus replication intermediates during the virus life cycle, which are calibrated as continuous deterministic ordinary differential equation-based models [4]. After calibration, these models are transformed into the discrete stochastic Markov chain-based models with time-inhomogeneous propensities that depend on the extracellular virus concentrations [1]. Stochastic intracellular models are simulated using the integral- or uniformization-based Monte Carlo algorithms, as well as the hybrid discrete-continuous stochastic-deterministic approximation methods [1, 2, 5, 6].

Results: The immune cell motility model calibrated to reproduce the lymphocyte movement patterns in lymphoid tissues was used to estimate the numbers of HIV-specific cytotoxic T cells required for effective control over the local bursts of HIV

infection in the lymph nodes [3]. The detailed models of intracellular virus replication were developed and calibrated for HIV-1, SARS-CoV-2 and LCMV [4–6]. Local and global sensitivity analysis was performed for deterministic intracellular models and for their discrete stochastic counterparts to identify the potential targets for antiviral drug development among the processes that regulate the virus life cycles. Markov chain-based models were used for quantifying the stochastic heterogeneity of intracellular virus replication and progeny production, depending on the multiplicity of cell infection. A multiscale model was developed to study the dynamics of the local bursts of HIV-infection in lymph nodes, integrating the intracellular, tissue, and systemic levels [1, 2]. The hybrid discrete-continuous stochastic-deterministic computational algorithm which is based on the uniformization Monte Carlo method was implemented to perform multiscale simulations of virus infection dynamics [1].

Conclusion: The essential element of the proposed multi-scale modelling approach is the use of hybrid discrete-continuous stochastic Markov chain-based formulations and simulation algorithms. This allows for prediction of intracellular heterogeneity or virus replication dynamics of the infected cells. Since target and immune cells are treated as individual agents and their positions are determined by the stochastic cell motility model, the associated stochasticity extends to the other scales. The developed computational tool provides a way to study virus infection dynamics at multiple scales of the immune system organization.

Funding: The study is supported by the Russian Science Foundation (grant No. 23-11-00116).

References

1. Grebennikov D.S. Computational methods for multiscale modelling of virus infection dynamics. *Russ J Numer Anal Math Modell*. 2023;38(2):75-87. doi 10.1515/rnam-2023-0007
2. Grebennikov D.S., Bocharov G.A. Spatially resolved modelling of immune responses following a multiscale approach: from computational implementation to quantitative predictions. *Russ J Numer Anal Math Modell.* 2019;34(5):253-260. doi 10.1515/rnam-2019-0021
3. Grebennikov D., Bouchnita A., Volpert V., Bessonov N., Meyerhans A., Bocharov G. Spatial lymphocyte dynamics in lymph nodes predicts the cytotoxic T cell frequency needed for HIV infection control. *Front Immunol*. 2019;10:01213. doi 10.3389/fimmu.2019.01213
4. Grebennikov D., Kholodareva E., Sazonov I., Karsonova A., Meyerhans A., Bocharov G. Intracellular Life Cycle Kinetics of SARS-CoV-2 Predicted Using Mathematical Modelling. *Viruses*. 2021;13(9):1735. doi 10.3390/v13091735
5. Sazonov I., Grebennikov D., Meyerhans A., Bocharov G. Markov Chain-Based Stochastic Modelling of HIV-1 Life Cycle in a CD4 T Cell. *Mathematics*. 2021;9(17):2025. doi 10.3390/math9172025
6. Sergeeva J., Grebennikov D., Casella V., Cebollada Rica P., Meyerhans A., Bocharov G. Mathematical Model Predicting the Kinetics of Intracellular LCMV Replication. *Mathematics*. 2023;11(21):4454. doi 10.3390/math11214454

doi 10.18699/bgrs2024-12.2-09

Complexity of immune suppression underlying immune privileges in stem system

Karpenko D.
National Medical Research Center for Hematology, Moscow, Russia
d_@list.ru

Key words: immune privileges; stem cells; stem system; immune suppression

Motivation and Aim: An ability to evade immune surveillance is called immune privileges. Immune privileges are recognized as a major issue and are wildly studied in context of cancer and cancer stem cells. Despite a lot less attention, there are studies demonstrated the immune privileges of non-pathologic stem cells in different tissues. We managed to demonstrate immune privileges of mesenchymal stem cells [1]. In the following review, I marked a deep mutual interference between immune and stem cells, and basing on the complexity of stem cells regulation suggested the term “stem system” [2]. Yet, the question about mechanisms underlying the phenomenon stayed unconsidered. I further continued to study literature for described molecular mechanisms responsible for immune suppression.
Methods and Algorithms: While the aim of the study is to aggregate data for non-pathologic cells and tissues, a significant part of data arises from studies of cancer or other pathologies. There are certain changes in pathologic regulation, it allows to identify associated molecular mechanisms. At the same time, there are mechanisms utilized by pathology, but not broken by itself. We can take examples from cancer, where tumor cells reprogram non-tumor endothelial or immune cells to support and protect tumor microenvironment.
Results: In addition to several mechanisms demonstrated to be involved in non-pathologic stem cells immune privileges, there are other mechanisms demonstrating contribution to immune suppression for pathologic conditions and for normal one. In general, immune suppression is activated by activation of immune cells to provide proper control and to prevent excessive damage. Immune suppression is performed by specialized immune cells and by other cells in various tissues, including stem cells. Many mechanisms are common for different cells, but could be differently regulated by side factors. Depending on the conditions, a signal can activate or suppress an immune cell. Conditions could be presented by a switching molecule or by quantitative values, as pH. Components of signal pathways are controlled by transcription, splicing and post transcription modifications. Variations in abundance of soluble and membrane bound signaling molecules can provide nonlinear response. Some receptor ligand relations still don’t identified. There are reports of new mechanisms involving microvesicles and contained in them long non-coding RNA and micro RNA. Important notice is that mechanisms are tangled with each other and other regulations. Partial duplications of studied mechanisms are common to provide robust control over immunity.
Conclusion: There are various immune suppressing mechanisms contributing to the complex network and to the formation of immune privileges. Mechanisms of immune suppression are required for the immune system to contribute to regeneration and to support homeostasis of stem cells. The stem system is also integrated in immune

regulation and may suppress the immune system. There are plenty of studies demonstrating individual mechanisms or their combination. The amount of available information makes aggregation a complex task requiring new automated approaches. The problem expands beyond a quantity of information to its complexity. As immune system is a protector from foreign invasions, restricting mechanisms could not afford to be simple. If an immune suppression would be achieved by simple regulation, it could be used by invading pathogens and so evolutionary discarded. On the contrary, immune suppressing mechanisms should be enigmatic as well as robust to protect and to support homeostasis. Deciphering immune regulation can provide solutions for medicine. Particular studies consider a number of factors, but there are always more unstudied factors to be involved. The cost of experimentally testing factor by factor grow faster than exponentially with the number of factors. It is worsened by differences between human and model animal. This results in the lack of knowledge forming a natural barrier to model on the aggregated data.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Karpenko D., Kapranov N., Bigildeev A. Nestin-GFP Transgene Labels Immunoprivileged Bone Marrow Mesenchymal Stem Cells in the Model of Ectopic Foci Formation. *Front Cell Dev Biol*. 2022;10:993056. doi 10.3389/FCELL.2022.993056/FULL
2. Karpenko D.V. Immune Privileges as a Result of Mutual Regulation of Immune and Stem Systems. *Biochemistry*. 2023;88(11):1818-1831. doi 10.1134/S0006297923110123

doi 10.18699/bgrs2024-12.2-10

Mathematical modeling of the promising composition of the next generation safe epitope vaccine against porcine salmonellosis

Lavrukhin M., Kichemazova N., Oglodina D., Larionova O., Zaitsev S., Feodorova V.*

Saratov State University of Genetics, Biotechnology and Engineering named after N.I. Vavilov, Saratov, Russia

** feodorovav@mail.ru*

Key words: *Salmonella*; immunity; *in silico* analysis; epitope; vaccine

Motivation and Aim: Currently, the methods of artificial neural network are considered to be one of the promising approaches for development of a new generation effective safe vaccines. The main advantages of these methods are a better prediction rate, speed, data efficiency and high rank of selection of MHC molecules which are capable to elicit successfully the protective immune response in target hosts. In fact, a rational design of either single- or multi-peptide vaccines predicted *in silico* could be critical to induce enhanced immunity against actual infection diseases in specific mammalian species. Moreover, these vaccines are absolutely safe because possess no live microbial material that highlights their great potency for both animal health and welfare, and overall, for global agricultural economy. *Salmonella enterica* subsp. *enterica* serovar Choleraesuis (*S. choleraesuis*), the causative agent of salmonellosis, is known as a swine-adapted foodborne pathogen associated to invasive infections in both animals and humans. Nowadays, *Salmonella* infection in humans is considered as a global health burden. Contaminated animal-derived foodstuffs, especially pork products are the main source for human salmonellosis. Overall, *Salmonella* infection in swine and other livestock could lead to marked economic losses in national pork industry and global agricultural sector as well.

Several live whole cell (LWC) and killed whole cell (KWC) vaccines for immunization against porcine salmonellosis are now available in Russian Federation and worldwide. However, all of them need to improve their level of immunogenicity and safety as well. The main goal of this study was to predict *in silico* potential ligands for CD8+ T-cell epitopes of MHC class I in pig model using freely available machine leaning algorithms.

Methods and Algorithms: The IEDB public database "T Cell Prediction – Class I" was used as the main tool for the current study. This tool is based on the selective simultaneous use of three different neural networks, such as: (i) conventional sparse encoding [1]; (ii) BLOSUM encoding [2]; and (iii) hidden Markov model encoding [3]. The NetMHCpan 4.1 EL was applied as the main predictive method by IEDB (recommended epitope prediction – 2023.09). Both nucleotide and amino acid sequences for 14 virulence genes (*ipfB*, *sopB*, *ompA*, *pipB*, *pipB2*, *sifA*, *sinH*, *sopD*, *sopD2*, *sopE2*, *sseB*, *spvB*, *spvC* and *spvR*) of the *S. choleraesuis* strain TC-177 were obtained from NCBI GenBank database (https://www.ncbi.nlm.nih.gov/nucleotide/). The relevant epitope data sets for MHC class I were predicted for all 44 available pig alleles (access date 02/24/2024). The panel of potentially protective epitopes was selected based on

3 main parameters, namely antigenicity, allergenicity and immunogenicity using the DDG (https://www.ddg-pharmfac.net/ddg/index.html) and IEDB database software (https://www.iedb.org), respectively, with default settings.

Results: Initially, in range 60,000–400,000 epitopes were predicted for each protein. Further, two consecutive cut-offs resulted for individual protein from 14 to 88 epitopes (0.007–0.044 %, i. e. lesser then 1 %) which were combined into 3 groups based on their desired immunological characteristics, such as: (i) the epitopes with marked antigenicity, immunogenicity and non-allergenic activity (n = in the range from 1 to 20), which may be involved in a protective immune response, and can become the basis for a new generation vaccine; (ii) the epitopes with marked antigenicity, immunogenicity and allergenic activity (n = in the range 1–13); and (iii) the largest group, including up to 90 % of epitopes (n = in the range 11–64) which demonstrated no required characteristics resulting no protective aberrant immune response in target animals. Importantly, the majority of potentially protective epitopes were associated with only 5 proteins encoded by such genes as *pipB*, *sinH*, *sopD*, *spvB* and *spvC*. In contrast, no epitopes with relevant protective activity were predicted for the proteins expressed by the remaining 9 genes (*ipfB*, *sopB*, *ompA*, *pipB2*, *sifA*, *sopD2*, *sopE2*, *sseB* and *spvR*).

Conclusion: The data obtained could be perspective for development of multi-epitope peptide vaccine against porcine salmonellosis.

Funding: The study is supported by Russian Science Foundation (No. 22-16-00165).

References

1. Baldi P., Brunak S. Bioiformatics: The Machine Learning Approach. A Bradford Book. Publisher: The MIT Press Cambridge, Massachusetts London, England, 2001
2. Henikoff S., Henikoff J.G. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA*. 1992;89(22):10915-10919. doi 10.1073/pnas.89.22.10915
3. Nielsen M., Lundegaard C., Worning P., Hvid C.S., Lamberth K., Buus S., Brunak S., Lund O. Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics.* 2004;20(9):1388-1397. doi 10.1093/bioinformatics/bth100

doi 10.18699/bgrs2024-12.2-11

A modular model of immune response as a computational platform to investigate a pathogenesis of infection disease

Miroshnichenko M.I.*, Kolpakov F.A., Akberdin I.R.

Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia

** Elloquse@gmail.com*

Key words: coronavirus; SARS-CoV-2; COVID-19; mathematical model; BioUML, immune response

Motivation and Aim: Since the initial applications of mathematical methods for modeling an immune response are dated in the 1970's [1], a multitude of models have emerged, covering all areas of immunology from autoimmune disorders to infectious diseases. Despite promising prospects, the field's evolution has been hampered by limited access to experimental data, insufficient methodology development, and the absence of suitable software to construct multicompartmental models with complex structure and functional interrelationships between biological entities. While some challenges have been effectively addressed, others, such as data availability, persist. The landscape dramatically shifted with the onset of the COVID-19 pandemic. The unprecedented scale and severity of the disease prompted international collaboration among numerous scientific groups, accelerating joint efforts to address this common challenge. The collective attempt yielded a lot of experimental data, significantly facilitating the validation process and contributing to the emergence of novel mathematical models for studying COVID-19 [2]. Taking into account some shortcomings of these models (a limited set of compartments, a small amount of used experimental data, and the absence of key components of the immune response), we aimed to develop a modular mathematical model of the SARS-CoV-2 infection process and the subsequent innate and adaptive immune responses. Our goal is not only to investigate the corresponding disease at a systemic level but also to provide a framework for further development and research in the field of host-pathogens (existing and future) interactions.

Methods and Algorithms: As a modeling tool, we employed our own open-source computational platform for systems biology, BioUML [3]. To build the modular model, we utilized the system of ordinary differential equations (ODEs) supplemented by the integration of delay differential equations (DDEs), which effectively capture the time-dependent dynamics of cell development and migration. Model simulations were conducted using the Java Variable-Coefficient ODE (JVODE) solver, which is integrated within the BioUML environment. Model validation was performed using experimental data of time-series measurements obtained from adult individuals with moderate COVID-19 symptoms, including viral load in the Upper Airways (UA) and Lungs, CD4+ and CD8+ T cells, and levels of key cytokines including IL-6, IL-12, and IFNγ. Following manual calibration, three optimization algorithms implemented in the BioUML platform [4–6], were applied to refine the model parameters. Then we selected the best scored one from a set of obtained parameter solutions. In addition, to evaluate the impact of different parameters and their variations on the progression of COVID-19, we performed a local sensitivity analysis also implemented in the BioUML platform.

Results: We have developed a multicompartmental model consisting of the Upper Airways, Lungs, and draining Lymph Nodes (Fig. 1). The Upper Airways represent a simplified version of the model, including only the coronavirus, epithelial cells in different states, and immature dendritic cells. The draining lymph nodes encompass CD8+ T and B cells, with the former differentiating into cytotoxic T cells responsible for eliminating infected epithelial cells. The latter differentiate into plasma cells that produce antibodies. In the lungs, which serve as the primary model, both macrophages and dendritic cells are included, reflecting the innate immunity component. In addition to CD8+ T cells, the lungs' draining lymph nodes also harbor CD4+ T cells, which differentiate into T helper type 1 and T follicular helper cells. Furthermore, the lymph nodes contain various types of cytokines, including IL-2, IL-6, IL-12, and IFNγ.

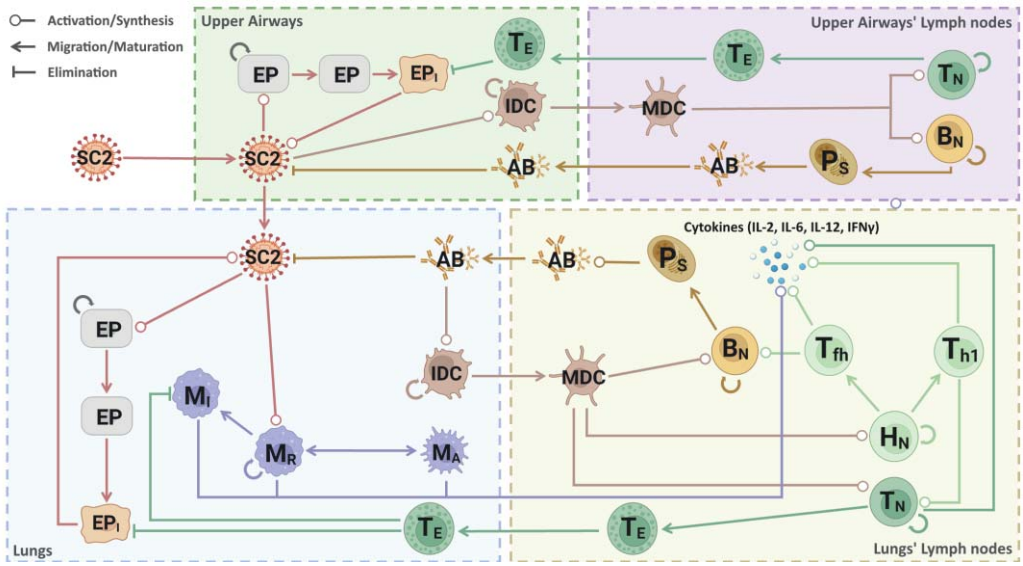


Fig. 1. A conceptual diagram of the model. SC2 – SARS-CoV-2, EP – Epithelial Cells, EP_I – Infected Epithelial Cells, M_R – Resting Macrophages, M_I – Infected Macrophages, M_A – Activated Macrophages, IDC – Immature Dendritic Cells, MDC – Mature Dendritic Cells, B_N – Naïve B Cells, T_N – Naïve CD8+ T Cells, H_N – Naïve CD4+ T Cells, Th1 – T Helper Type 1 Cells, Tfh – T Follicular Helper Cells, T_E – Cytotoxic T Cells, P_S – Plasma Cells, AB – Antibodies

After fitting the model to experimental quantitative data, we reproduced the main disease severity scenarios: moderate, severe, and critical (Fig. 2). We used the level of IL-6 as an indicator of the severity extent, as it most strongly correlates with worse outcomes. In addition to this, we considered the degree of epithelial cell damage and viral load. To model severe disease, we adjusted parameters associated with aging-related alterations in the immune system, given that age is a major risk factor for deteriorating the COVID-19 outcome [7]. Additionally, we investigated the potential differences among SARS-CoV-2 variants in terms of infectivity, transmissibility, and their ability to evade the immune response. Furthermore, we modeled the immune response to the infection in patients with HIV and demonstrated how interferon treatment affects the course of the disease.

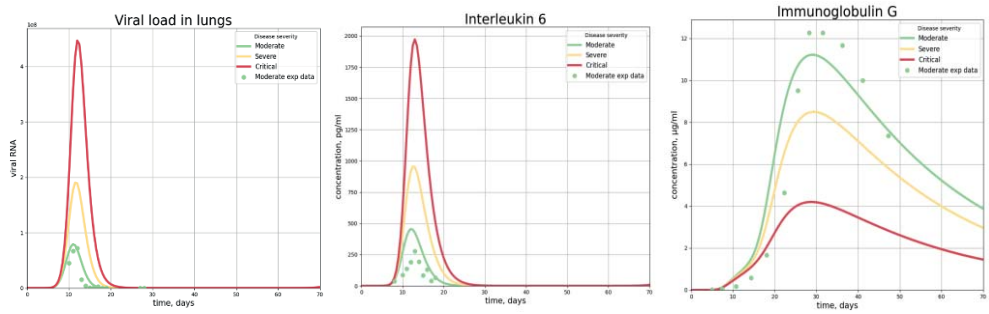


Fig. 2. The modular model simulations of viral load, IL-6 and IgG for different COVID-19 severity scenarios. Green line represents moderate progression, yellow line designates severe progression and red line indicates critical progression according to the model simulations, while dots represent experimental measurements for corresponding variables

Conclusion: Consequently, the developed modular model encompasses 31 ODEs and 114 parameters. It not only simulates diverse disease scenarios, including moderate or severe progression, but also is able to reproduce the co-infection with HIV, which implies different extent of immunosuppression. Furthermore, it simulates primary changes in SARS-CoV-2 variants, such as differences between Omicron and other strains as well as various treatment strategies. It is worth to note that constructed modules of the model can be employed as "building blocks" for simulating other infectious diseases taking into account follow-up immune response. The modular model and all experimental data are available at gitlab: https://gitlab.sirius-web.org/diploms/modular-immune-system.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation, (Agreement 075-10-2021-093, Project CMB-RND-2123: "Virtual patient").

References

1. Jílek M. On A Generalized Mathematical Model of the Immune Response. *Acta. Appl. Math.* 1989;14:115-123. doi 10.1007/978-94-009-2358-4_11
2. Frazer N., Brierley L. et al. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLoS Biol*. 2021;19(4). doi 10.1371/journal.pbio.3000959
3. Kolpakov F., Akberdin I. et al. BioUML – towards a universal research platform. *Nucleic Acids Res*. 2022;50(1):124-131. doi 10.1093/nar/gkac286
4. Runarsson T. P., Yao X. Stochastic ranking for constrained evolutionary optimization. *IEEE Transactions on Evolutionary Computation*. 2000;4(3):284-294. doi 10.1109/4235.873238
5. Kennedy J., Eberhart R. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*. 1995;4:1942-1948. doi 10.1109/ICNN.1995.488968
6. Nebro A. J., Durillo J. J. et al. MOCell: A cellular genetic algorithm for multiobjective optimization. *International Journal of Intelligent Systems*. 2009;24(7):726-746. doi 10.1002/int.20358
7. Biswas M., Rahaman S. et al. Association of Sex, Age, and Comorbidities with Mortality in COVID-19 Patients: A Systematic Review and Meta-Analysis. *Intervirology*. 2021;64(1):36-47. doi 10.1159/000512592

doi 10.18699/bgrs2024-12.2-12

A role for CD4 helper cells in HIV control and progression

Rouzine I.M.*

Sechenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences, Saint Petersburg, Russia

** igor.rouzine@iephb.ru*

Key words: mathematical modeling; post-treatment control

Motivation and Aim: Untreated HIV infection is characterized by persistent viremia and a gradual loss of CD4 T cells eventually causing AIDS [1, 2]. However, a small fraction of individuals is capable of controlling HIV replication at low levels in the absence of antiretroviral treatment (ART) [3, 4]. The other individuals, termed "post-treatment controllers", can achieve control after an early antiviral therapy [5] or antibody treatment [6]. Controllers progress to AIDS very slowly or not at all [3, 4]. Substantial evidence demonstrates that controllers suppress HIV replication due to an active immune response including cytotoxic CD8 T cells and CD4 T cells [7, 8]. The range of the functional avidity of HIV-specific CD8 T cells does not differ between a progressive and non-progressive infection [9]. A striking feature setting the two cohorts apart, which remains unexplained, is that an active CD4 T cell response with IL-2 secretion capacity and proliferative responses is observed in controllers but not in progressors [10]. The CD4 T cell avidity distribution for progressors is much narrower than for controllers and centered at much lower values (a higher threshold in antigen) (Fig. 1) [11].

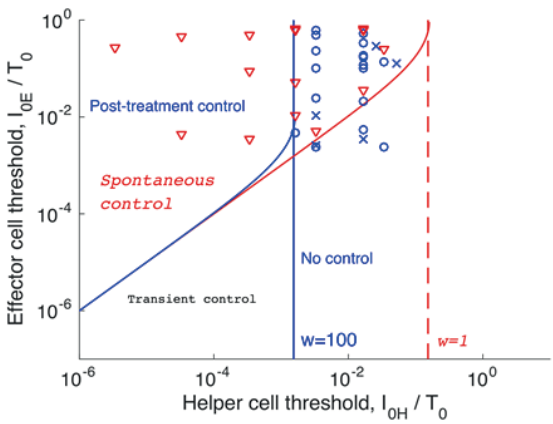


Fig. 1. The difference in CD4 cell avidity between controllers and progressors based on [11]. X-axis: Rescaled inverse functional avidity of CD4 T cells, denoted I_{H0}, measured in three different cohorts of patients [11]. Y-axis: simulated points of the inverse avidity of CD8 T cells, denoted I_{E0}, distributed in a way similar to data from patients in another study [9]. Red symbols (o) correspond to spontaneous controllers, and blue symbols (o, *) correspond to untreated patients with high viremia and patients on long-term ART, respectively. Blue and red lines show the phase diagrams predicted in this work at two different values of infectivity ratio between activated and resting CD4 T cells, $w = 100$ and $w = 1$. The other parameters are given in [12] The methods of measurement and the units are different between the CD8 and CD4 avidity, so that X-axis and Y-axis use two separate scaling factors found from fitting [12]

The present work offers a model that helps to understand this puzzle, as well as explains the differences in viral dynamics between controllers and progressors, with potential clinical applications.

Methods and Algorithms: The key idea is that the survival strategy of HIV may be to infect and suppress CD4 T cells and thus impair the immune system by making it much less sensitive to the presence of the virus. The model (Fig. 2) comprises five cell compartments. To elucidate the systems dynamics, two parallel methods are used: mathematical analysis and numeric computer simulation. It is formalized by a system of differential equations

$$\frac{dT}{dt} = b - pIT - d_T T,\ \frac{dI}{dt} = [1 - p_L(E)]pIT - d_I\left(1 + \frac{E}{E_0}\right)I + r(E)L$$

$$\frac{dL}{dt} = p_L(E)pIT - r(E)L,\ \frac{dH}{dt} = cH\alpha\left(\frac{I}{I_{0H}}\right)\beta\left(\frac{H}{H_m}\right) - pwIH - d_H H$$

$$\frac{dE}{dt} = cE\alpha\left[\frac{I}{I_{0E}} + \frac{H}{H_0}\alpha\left(\frac{I}{I_{0H}}\right)\right]\beta\left(\frac{E}{E_m}\right) - d_E E,\ r(E) = r_0 + r_{max}\frac{E}{E+E_{0L}},\ p_L(E) = p_{L0}\frac{E_{0L}}{E+E_{0L}}$$

$$\alpha(x) \equiv 1 - e^{-x}, \beta(y) \equiv (1-y)^{\frac{1}{3}}, p = 0 \text{ if } I < \frac{I_{in}}{2}$$

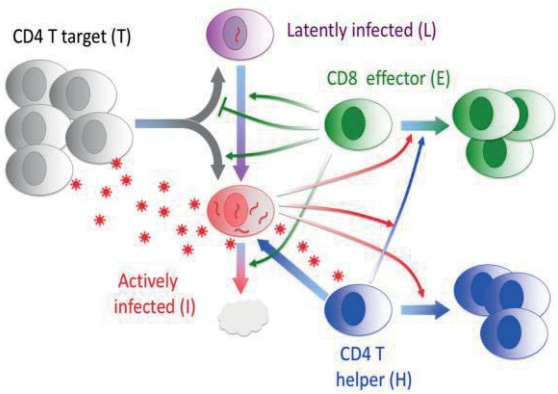


Fig. 2. The model of the immune response and virus dynamics. Included are five cell compartments, as follows: uninfected but susceptible target cells of CD4 T CCR5 phenotype, which number is denoted T, actively infected cells, I, latently-infected cell subset, L [16, 17], virus-specific helper CD4 T cells, H, and cytotoxic effector CD8 T cells, E. Broad arrows show cell flow and proliferation, thin arrows show control signals. Little red stars are free virions. Based on [12]

The general ranges of parameters are taken from [13, 14]. The specific (default) values of parameters are adjusted to fit data in Fig. 1, to obtain oscillation-free dynamics, and to match the representative levels of various cell types (see experimental references in [15]).

Results: Numeric simulation and analysis of the model above demonstrates that the system has two possible stable steady states. The helper-dependent steady state has a very low virus load corresponding to a spontaneous or post-treatment controller. The helper-independent state has no helper cells and a high virus load. If the ratio of infectivity of virus in activated HIV-specific to resting non-HIV specific cells w is on the order of 1, and the avidity of helper cells is sufficiently high, a controller state is established spontaneously, consistent with the clinical observation [11].

If parameter w is much larger than 1, helper cells are completely depleted during the acute phase of infection before they have a chance to reach their stable steady-state level (Fig. 3A). As a result, the high-viremia state ensues indefinitely (until AIDS not considered here). The low viremia state can be rescued with the help of early external intervention, such as anti-retroviral therapy [5] or broadly neutralizing antibodies [6]. The model explains this effect: by blocking infection of helper cells, ART started early rescues helper cells (Fig. 3B). ART must start in a narrow window depending on patient's parameters: basic reproduction ratio R_0, the initial level of naive T cells, and cell lifespans. Data for CD4 T cell avidity obtained for different patient cohorts [11] agree with the phase diagrams predicted by the model for two different values of the infectivity asymmetry parameter (Fig. 1). Because w is typically quite high [13, 14], post-treatment control is predicted to be much more likely than the spontaneous control, which agrees with observations (10–15 and 0.5% of patients, respectively).

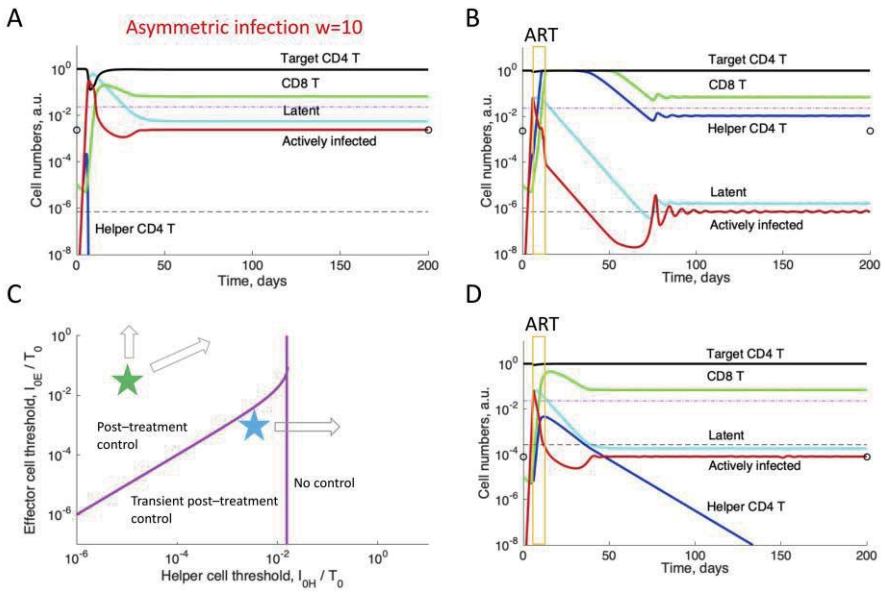


Fig. 3. A much higher infectivity of activated HIV-specific than resting non-specific CD4 T cells predicts a post-treatment controller. Infectivity of virus-specific (and hence activated) CD4 T cells is assumed to be 10 times higher than for non-specific (and hence resting) CD4 T cells, $w = 10$. (A) Early infection and killing of CD4 T cells leads to a high-viremia helper-free state. (B) ART between days 6 and 12 with the efficacy of 97% leads to the rescue of helper cells and establishment of a low-viremia state. (D) If avidity of helper cells is lower than the avidity of killer cells, helper cells decay gradually and control is only transient. (C) Phase diagram predicting the three outcomes of infection. The asterisks show parameters used in simulation: (A,B) $I_{0E} = 0.03, I_{0H} = 9 \cdot 10^{-6}$, (D) $I_{0E} = 0.001, I_{0H} = 0.003$. Relevant parameters in (D) are $w = 10, T_0/[R_0 d_I \int I(t)dt] = 0.2$. Based on [12]

The model does not include the "exhaustion" of CD8 T cells previously proposed to explain the divergent outcome in controllers and progressors [18]. In fact, CD8 T cell anergy is associated with long-term chronic infections [19, 20], while the outcome of HIV infection is decided during the first weeks. Anergy is usually modeled using either a time integral of virus load [21] or a third-party cell type [22]. More likely, exhaustion is not the cause but a consequence of persistence, which occurs for an unrelated reason. Also, the exhaustion model does not use any virological details specific for HIV and does not explain why all viruses do not persist like HIV. The phase diagram in Fig. 3C differs substantially from the diagram predicted by "exhaustion" model [18]. In particular, it does not depend on the killing efficacy of CTLs.

Conclusion: The study [12] offers a model that predicts a testable connection between HIV control, T cell avidity, and the dependence of HIV infectivity on cell activation status that agrees with data better than the previous "exhaustion" model. It can be used to more exactly time the time of ART to achieve post treatment control. The rather strict timing requirement raises the possibility that the currently-observed post-treatment controllers may be only a small fraction of their potential number.

Funding: The study was carried out within the framework of the governmental assignment to the IEPhB RAS (No. 075-00264-24-00).

References

1. Okoye A.A., Picker L.J. CD4(+) T-cell depletion in HIV infection: mechanisms of immunological failure. *Immunol Rev.* 2013;254(1):54-64. doi 10.1111/imr.12066
2. Rouzine I.M. An evolutionary model of progression to AIDS. *Microorganisms*. 2020;8(11):1714. doi 10.3390/microorganisms8111714
3. Arnaout R.A., Lloyd A.L., O'Brien T.R., Goedert J.J., Leonard J.M., Nowak M.A. A simple relationship between viral load and survival time in HIV-1 infection. *Proc Natl Acad Sci USA.* 1999;96(20):11549-11553. doi 10.1073/pnas.96.20.11549
4. Lambotte O., Boufassa F., Madec Y. et al. HIV controllers: a homogeneous group of HIV-1-infected patients with spontaneous control of viral replication. *Clin Infect Dis*. 2005;41(7):1053-1056. doi 10.1086/433188
5. Samri A., Bacchus-Souffan C., Hocqueloux L. et al. Polyfunctional HIV-specific T cells in Post-Treatment Controllers. *AIDS*. 2016;30(15):2299-2302. doi 10.1097/QAD.0000000000001195
6. Nishimura Y., Gautam R., Chun T.W. et al. Early antibody therapy can induce long-lasting immunity to SHIV. *Nature.* 2017;543(7646):559-563. doi 10.1038/nature21435
7. Saez-Cirion A., Lacabaratz C., Lambotte O. et al. HIV controllers exhibit potent CD8 T cell capacity to suppress HIV infection ex vivo and peculiar cytotoxic T lymphocyte activation phenotype. *Proc Natl Acad Sci USA*. 2007;104(16):6776-6781. doi 10.1073/pnas.0611244104
8. Pitcher C.J., Quittner C., Peterson D.M. et al. HIV-1-specific CD4+ T cells are detectable in most individuals with active HIV-1 infection, but decline with prolonged viral suppression. *Nat Med.* 1999;5(5):518-525. doi 10.1038/8400
9. Vigano S., Bellutti Enders F., Miconnet I. et al. Rapid perturbation in viremia levels drives increases in functional avidity of HIV-specific CD8 T cells. *PLoS Pathog.* 2013;9(7):e1003423. doi 10.1371/journal.ppat.1003423
10. Potter S.J., Lacabaratz C., Lambotte O. et al. Preserved central memory and activated effector memory CD4+ T-cell subsets in human immunodeficiency virus controllers: an ANRS EP36 study. *J Virol.* 2007;81(24):13904-13915. doi 10.1128/JVI.01401-07
11. Vingert B., Perez-Patrigeon S., Jeannin P. et al. HIV controller CD4+ T cells respond to minimal amounts of Gag antigen due to high TCR avidity. *PLoS Pathog*. 2010;6(2):e1000780. doi 10.1371/journal.ppat.1000780
12. Rouzine I.M. A role for CD4 + helper cells in HIV control and progression. *AIDS*. 2022;36(11):1501-1510. doi 10.1097/QAD.0000000000003296
13. Sergeev R.A., Batorsky R.E., Coffin J.M., Rouzine I.M. Interpreting the effect of vaccination on steady state infection in animals challenged with Simian immunodeficiency virus. *J Theor Biol.* 2010;263(3):385-392. doi 10.1016/j.jtbi.2009.12.018

14. Sergeev R.A., Batorsky R.E., Rouzine I.M. Model with two types of CTL regulation and experiments on CTL dynamics. *J Theor Biol.* 2010;263(3):369-384. doi 10.1016/j.jtbi.2009.11.003
15. Rouzine I.M., Weinberger A.D., Weinberger L.S. An evolutionary role for HIV latency in enhancing viral transmission. *Cell.* 2015;160(5):1002-1012. doi 10.1016/j.cell.2015.02.017
16. Chun T.W., Engel D., Berrey M.M., Shea T., Corey L., Fauci A.S. Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proc Natl Acad Sci USA.* 1998;95(15):8869-8873. doi 10.1073/pnas.95.15.8869
17. Sedaghat A.R., Dinoso J.B., Shen L., Wilke C.O., Siliciano R.F. Decay dynamics of HIV-1 depend on the inhibited stages of the viral life cycle. *Proc Natl Acad Sci USA.* 2008;105(12):4832-4837. doi 10.1073/pnas.0711372105
18. Conway J.M., Perelson A.S. Post-treatment control of HIV infection. *Proc Natl Acad Sci USA.* 2015;112(17):5467-5472. doi 10.1073/pnas.1419162112
19. Khaitan A., Unutmaz D. Revisiting immune exhaustion during HIV infection. *Curr HIV/AIDS Rep.* 2011;8(1):4-11. doi 10.1007/s11904-010-0066-0
20. Zajac A.J., Blattman J.N., Murali-Krishna K. et al. Viral immune evasion due to persistence of activated T cells without effector function. *J Exp Med.* 1998;188(12):2205-2213
21. Bocharov G., Klenerman P., Ehl S. Modeling the dynamics of LCMV infection in mice: II. Compartmental structure and immunopathology. *J Theor Biol.* 2003;221:349-378
22. Rouzine I., Murali-Krishna K., Ahmed R. Generals die in friendly fire, or modeling immune response to HIV. *J Appl Comp Math.* 2005;184:258-274

doi 10.18699/bgrs2024-12.2-13

Modeling of the immunological process of HIV infection development at an early stage

Surnin P.[1]*, Shishlenin M.[1], Bocharov G.[2]

[1] *S.L. Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia*

[2] *G.I. Marchuk Institute of Computational Mathematics, RAS, Moscow, Russia*

* *p.surnin@internet.ru*

Key words: HIV; parameter identification; parameter sensitivity analysis

Motivation and Aim: The human immunodeficiency virus (HIV) attacks the immune system and weakens protection against many infections. So far, there are no methods to achieve complete elimination of HIV from the body of an infected person. However, due to increased access to HIV prevention, diagnosis and treatment with the help of HAART, HIV infection has moved into the category of controlled chronic diseases. Mathematical modeling methods are actively used to study the kinetic mechanisms of the pathogenesis of HIV infection and the development of personalized treatment approaches based on combined immunotherapy. One of the central tasks of HIV infection modeling is to determine the individual response parameters of the immune system in the acute phase of HIV infection.

Methods and Algorithms: A system of ordinary differential equations is considered [1]. The model consists of eight equations describing four states of CD4+ T cells, two types of CD8+ T cells that relate to human cellular immunity, and two equations for viral load. The Sobol method was used for sensitivity analysis. The search for the optimal value of the parameters was performed using a stochastic optimization algorithm.

Results Taking into account the sensitivity analysis performed by the Sobol method, the inverse problem was solved for the parameters of the model, the reconstructed solution of which describes the dynamics of the acute phase of HIV infection according to clinical data [2].

Funding: The work is supported by the Mathematical Center in Akademgorodok under the agreement No. 075-15-2022-281 with the Ministry of Science and Higher Education of the Russian Federation.

References

1. Banks H.T., Davidian M., Hu Sh., Kepler G.M., Rosen-berg E.S. Modelling HIV immune response and validation with clinical data. *Journal of Biological Dynamics*. 2008;2(4):357-385
2. Kazer S.W., Aicher T.P., Muema D.M. et al. Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. *Nat Med*. 2020;26:511-518. doi 10.1038/s41591-020-0799-2

12 Симпозиум «Математические проблемы биоинформатики и системной компьютерной биологии. Анализ больших генетических данных и искусственный интеллект»

Symposium “Mathematical problems of bioinformatics and systems computational biology. Big genetic data analysis and artificial intelligence”

doi 10.18699/bgrs2024-12.3-01

Применение таргетированного метаболомного скрининга клеток методом ВЭЖХ-МС/МС в изучении биологических эффектов ТГц-излучения

Басов Н.В.[1, 2], Бутикова Е.А.[1, 3]*, Рогачев А.Д.[1, 2], Сотникова М.А.[1], Гайслер Е.В.[1], Разумов И.А.[1, 3, 4], Соловьева О.И.[1, 3, 4], Сотникова Ю.С.[1, 2, 5], Патрушев Ю.В.[1, 5], Коломеец Д.А.[3], Каныгин В.В.[1, 3], Попик В.М.[3], Покровский А.Г.[1]

[1] *Новосибирский государственный университет, Новосибирск, Россия*

[2] *Новосибирский институт органической химии им. Н.Н. Ворожцова СО РАН, Новосибирск, Россия*

[3] *Институт ядерной физики им. Г.И. Будкера СО РАН, Новосибирск, Россия*

[4] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[5] *Институт катализа им. Г.К. Борескова СО РАН, Новосибирск, Россия*

* *katabutikova@gmail.com*

Ключевые слова: пробоподготовка; метаболомика; пробоподготовка клеток; ТГц; метаболомика клеток

Мотивация и цель: Метаболомика – концепция в биоаналитической химии, целью которой является количественная оценка набора малых молекул (с молекулярной массой менее 1.5 кДа), присутствующих в биологической системе в любом из ее физиологических состояний. Таргетированный скрининг метаболитов методом высокоэффективной жидкостной хроматографии с масс-спектрометрической детекцией (ВЭЖХ-МС/МС) может служить удобным и чувствительным инструментом для выявления ранних изменений в живых системах на молекулярном уровне. Ранее нами был разработан подход к метаболомному скринингу методом ВЭЖХ-МС/МС с использованием монолитной колонки на основе 1-винил-1,2,4-триазола, который позволил детектировать более 400 метаболитов за один анализ [1].

Терагерцовое излучение (терагерцовые волны, субмиллиметровое излучение, Т-лучи, Т-волны, ТГц) находится между микроволновой и инфракрасной областями электромагнитного спектра, диапазон которого определятся от частот 100 ГГц до 10 ТГц [2]. Изучение потенциального воздействия Т-волн на биологические системы востребовано для актуализации протоколов безопасности ТГц и разработки большого числа его практических приложений. Биологические эффекты ТГц-излучения сложны в определении, а метаболомный подход применим в установлении тонких биохимических изменений, индуцированных электромагнитным излучением.

Целью работы являлась разработка подхода к пробоподготовке и таргетированному метаболомному скринингу клеток для поиска биохимических изменений в клетках в ответ на ТГц-излучение.

Методы и алгоритмы: В данном исследовании было проведено облучение клеток линии SK-MEL-28 (модель меланомы человека) на Новосибирском лазере на свободных электронах (НЛСЭ) [3] с параметрами: 2.3 ТГц, длина волны 130 мкм

в течение 10 и 45 минут. Клеточные образцы анализировали ВЭЖХ-МС/МС в режимах гидрофильной и обращенно-фазовой хроматографии.
Результаты. Было показано, что в биологических образцах, полученных из 10 тысяч клеток, можно обнаружить ряд полярных метаболитов и липидов, однако наибольшее количество обнаруженных метаболитов достигается при анализе образцов от 500 тысяч клеток и выше.
Подход к анализу клеточных образцов был применен для метаболомного скрининга клеток, облученных ТГц на НЛСЭ «Изучение воздействия ТГц излучения на живые системы». Метаболомный анализ позволил получить данные о содержании 407 метаболитов, из которых 40 значимо отличались между группами облученных ТГц клеток и контрольной группой. Наблюдались изменения в содержании метаболитов, включая пути метаболизма пуриновых и пиримидиновых азотистых оснований. Наиболее значительные изменения отмечены в пуриновом метаболизме, где наблюдалось систематическое снижение содержания метаболитов, таких как АТФ, АДФ и АМФ, в то время как содержание продуктов деградации нуклеотидов, таких как ксантин и гипоксантин, увеличивалось. Значительные изменения также выявлены в пиримидиновом метаболизме, где как при 10-минутном, так и при 45-минутном облучении отмечено повышение конечных метаболитов, таких как уридин и урацил. Это первое исследование, где применяется метаболомный скрининг в биологических образцах, облученных ТГц, которое демонстрирует ТГц-индуцированные метаболические изменения в клетках человека. Полученные результаты могут свидетельствовать о нарушении функций энергетических систем клеток.
Выводы: Результаты работы показывают, что облучение клеток меланомы ТГц приводит к комплексным изменениям в их метаболическом профиле. Эти изменения включают снижение содержания ключевых метаболитов энергетического обмена, таких как АТФ, и увеличение продуктов их деградации. Отмечено, что разработанный нами подход позволяет получать наиболее полную картину метаболомного профиля при анализе от 500 тысяч клеток в образце, что может быть использовано при изучении экзогенных эффектов на редких клетках. Исходя из наших результатов, метаболомика является эффективным инструментом для изучения механизмов действия терагерцового излучения на клеточном уровне, а результаты метаболомного скрининга могут быть полезны в понимании биологических эффектов терагерцового излучения и разработке стратегий безопасности его применения в различных областях.
Финансирование: Работа выполнена на уникальной установке Новосибирский лазер на свободных электронах на базе ЦКП «Сибирский центр синхротронного и терагерцового излучения» при поддержке гранта РНФ (Проект No. 19-72-202). Исследование клеток поддержано средствами государственного задания БЧ 2020-0039-1. Анализ методом ВЭЖХ-МС/МС поддержан средствами государственного задания No. FSUS-2020-0035. Изготовление монолитных колонок для HPLC поддержано средствами проекта No. FWUR-2024-0032.

The use of targeted metabolomic cell screening by LC-MS/MS in the study of the effects of THz radiation

Basov N.V.[1, 2], Butikova E.A.[1,3]*, Rogachev A.D.[1, 2], Sotnikova M.A.[1], Gaisler E.V.[1], Razumov I.A.[1, 3, 4], Solovieva O.I.[1, 3, 4], Sotnikova Y.S.[1, 2, 5], Patrushev Y.V.[1, 5], Kolomeyets D.A.[3], Kanygin V.V.[1, 3], Popik V. M.[3], Pokrovsky A.G.[1]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry SB RAS, Novosibirsk, Russia*

[3] *Budker Institute of Nuclear Physics SB RAS, Novosibirsk, Russia*

[4] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[5] *Boreskov Institute of Catalysis, Novosibirsk, Russia*

* *katabutikova@gmail.com*

Key words: sample preparation; metabolomics; preparation of cells; THz; cell metabolomics

Motivation and Aim: Metabolomics is a branch of bioanalytical chemistry that aims to quantify small molecules (with a molecular weight less than 1.5 kDa) in biological systems in various physiological states. High-performance liquid chromatography with mass spectrometry detection (LC-MS/MS) is a powerful tool for targeted screening of metabolites, allowing for the detection of early changes at the molecular level in living systems. Previously, we developed a method for metabolomic screening using LC-MS/MS with a monolithic column made from 1-vinyl-1,2,4-triazole [1]. This approach allowed us to detect more than 400 different metabolites in a single analysis. Terahertz radiation, also known as terahertz waves, submillimeter radiation, T-rays or THz, is located between the microwave and infrared regions of the electromagnetic spectrum. This range is determined by frequencies between 100 GHz and 10 THz [2]. The study of potential effects of this radiation on biological systems is important for updating safety protocols and developing practical applications. However, it is difficult to determine the biological effects of terahertz radiation due to its subtle nature. The metabolomic approach can be used to identify biochemical changes caused by electromagnetic radiation. The goal of this work was to develop a method for sample preparation and targeted metabolomics screening of cells in order to identify biochemical changes in response to terahertz radiation.

Methods and Algorithms: In this study, the cells of the SK-MEL-28 line (human melanoma model) were irradiated using the Novosibirsk Free Electron Laser (NovoFEL) [3] with parameters: 2.3 THz, wavelength 130 μm for 10 and 45 minutes. Cell samples were analyzed by LC-MS/MS in hydrophilic and reverse-phase chromatography modes.

Results: It has been shown that a number of polar metabolites and lipids can be detected in biological samples obtained from 10,000 cells or more. However, the largest number of metabolites detected is achieved when analyzing samples from 500,000 or more cells. The approach to the analysis of cell samples was used for the metabolomic screening of cells exposed to THz radiation in the NovoFEL study “Investigating the effects of THz radiation on living systems”. Metabolomic analysis allowed us to obtain data on the content of 407 metabolites, 40 of which differed significantly between the groups of THz-irradiated cells and the control group. Changes in the content of metabolites were observed, including in the metabolic pathways of purine and pyrimidine nitrogenous bases. The most significant changes occurred in purine metabolism, where a systematic

decrease in the content of ATP, ADP, and AMP was noted, while the levels of nucleotide degradation products, such as xanthine and hypoxanthine, increased. Significant changes were also found in pyrimidine metabolism. Both 10- and 45-minute irradiations showed an increase in the levels of end metabolites, such as uridine and uracil. This is the first study to use metabolomics screening in biological samples exposed to THz radiation, which demonstrates THz-induced metabolic changes in human cells. The results obtained suggest that the energy systems of these cells may be disrupted.

Conclusion: The results show that the irradiation of melanoma cells with THC leads to complex changes in their metabolic profile. These changes include a decrease in the levels of key metabolites involved in energy metabolism, such as ATP, and an increase in the levels of their degradation products. It is noted that our developed approach allows us to obtain a more complete picture of the metabolomic profile by analyzing 500,000 cells in a sample. This can be used to study the effects of exogenous factors on rare cells. Based on our findings, metabolomics is an effective tool for investigating the mechanisms of terahertz radiation action at the cellular level. The results of metabolomic screening can provide valuable insights into the biological effects of this radiation and help develop safety strategies for its use in various fields.

Funding: The experimental part of the study was carried out with the support of the Russian Science Foundation (Project No. 19-72-202) at the unique Novosibirsk Free Electron Laser installation using equipment from the Siberian Center for Synchrotron and Terahertz Radiation. The funds of the state task number BC 2020-0039-1 were involved in the work. LC-MS/MS analysis has been conducted with the support of a state project No. FSUS-2020-0035. The preparation of monolithic columns for HPLC has been conducted with the support of a state project No. FWUR-2024-0032.

Список литературы/References

1. Basov N.V., Rogachev A.D. et al. Global LC-MS/MS targeted metabolomics using a combination of HILIC and RP LC separation modes on an organic monolithic column based on 1-vinyl-1, 2, 4-triazole. *Talanta*. 2024;267:125168. doi 10.1016/j.talanta.2023.125168
2. Ghann W., Uddin J. Terahertz (THz) spectroscopy: a cutting-edge technology. *ed. by J. Uddin. USA: Coppin State University*. 2017;1:3-20. doi 10.5772/67031
3. Kulipanov G.N. et al. Novosibirsk free electron laser-facility description and recent experiments. *IEEE Trans Terahertz Sci Technol*. 2015;5(5):798-809. doi 10.1109/TTHZ.2015.2453121

doi 10.18699/bgrs2024-12.3-02

Современные программные решения для преобразования геномных координат между различными версиями сборок генома человека

Бобрик П.Ю.*, Вертёлко В.Р., Гурьянова И.Е.

Республиканский научно-практический центр детской онкологии, гематологии и иммунологии, Боровляны, Республика Беларусь

** bobrykpavel2001@gmail.com*

Ключевые слова: преобразование геномных координат; liftOver; GRCh37; GRCh38

Мотивация и цель: Референсная геномная сборка является необходимым ресурсом при работе с данными высокопроизводительного секвенирования. На сегодняшний день существуют две актуальные версии геномных сборок человека – GRCh37 и GRCh38. Самой свежей считается версия GRCh38, одним из основных обновлений в которой стало закрытие многочисленных пробелов, присутствующих в предыдущей сборке [1]. Несмотря на это, многие исследователи не решаются перейти на актуальную версию геномной сборки, так как подобный переход может потребовать значительных изменений в алгоритме обработки данных: установка/настройка дополнительных программных решений, скачивание/интеграция новых баз данных, а также редактирование общего пайплайна обработки. В свою очередь, использование GRCh37 накладывает ряд ограничений на возможность работы с актуальными базами данных, так как их авторы сфокусировались на свежей сборке GRCh38. Если рассмотреть задачу поиска вариантов, то, в случае наличия сырых FASTQ-файлов, можно повторно провести процедуру маппинга и вызова вариантов для актуальной GRCh38. Данный процесс может потребовать дополнительных вычислительных и временных ресурсов, что зачастую не всегда возможно. Как альтернативу данному решению можно рассмотреть наиболее эффективную в плане используемых ресурсов процедуру, а именно – конвертацию геномных координат. Учитывая все вышесказанное, а также то, что наиболее популярный инструмент для преобразования геномных координат NCBI's Remap Tool прекратил свою работу в ноябре 2023 г., была поставлена цель провести обзор существующих программных решений, имеющих схожий функционал.

Методы и алгоритмы: Были протестированы следующие биоинформатические инструменты для преобразования геномных координат: CrossMap, LiftoverVcf (Picard), transanno, Assembly Converter (Ensembl). Для оценки работоспособности программ использовалось 10 VCF-файлов (≈1.5 Гб каждый), содержащих в среднем 4.5 млн замен, выявленных в ходе полногеномного секвенирования, а также файл GRCh37_to_GRCh38.chain, необходимый для выполнения конвертации. Тестирование осуществлялось в ОС Ubuntu (22.04).

Результаты: Программа CrossMap смогла обработать до 98,70 % всех замен в среднем за 2 минуты 17 секунд (±10 с) с минимальным использованием вычислительных ресурсов. Инструмент transanno показал схожие результаты: было преобразовано 98.72 % вариантов за 2 минуты 7 секунд (±6 с). Программа

LiftoverVcf (Picard) конвертировала 98.67 % замен в среднем за 1 минуту 15 секунд (±5 с), однако потребление оперативной памяти в пиковый момент достигало 12.8 Гб. Веб-инструмент Assembly Converter (Ensembl) дает возможность запустить преобразование координат в удобном графическом интерфейсе. Стоит отметить, что максимальный размер загружаемого в Assembly Converter файла не должен превышать 50 Мб, что накладывает ограничение на работу с данными полноэкзомного и полногеномного секвенирования.
Выводы: Было рассмотрено четыре инструмента для преобразования геномных координат. CrossMap и transanno можно назвать наиболее универсальными вариантами для данной задачи, так как данные программы легко устанавливаются и не требуют больших вычислительных мощностей. Assembly Converter является хорошим вариантом для специалистов с небольшим опытом работы в командной строке. LiftoverVcf (Picard) удобен в случаях, когда нужно провести преобразование в более сжатые сроки за счет дополнительных вычислительных ресурсов.

Modern software solutions for converting genomic coordinates between different versions of human genome assemblies

Bobryk P.Y.*, Vertelko V.R., Guryanova I.E.
Belarusian Research Center for Pediatric Oncology, Hematology and Immunology, Borovlyany, Belarus
** bobrykpavel2001@gmail.com*

Key words: genomic coordinates transformation; liftOver; GRCh37; GRCh38

Motivation and Aim: The reference genome assembly is a necessary resource for processing high-throughput sequencing data. Today, there are two current versions of human genome assemblies, GRCh37 and GRCh38. The newest version is GRCh38, with one of the major updates being the closure of numerous gaps present in the previous assembly [1]. Despite this, many researchers hesitate to switch to the current version of the genomic assembly, as such a switch may require significant changes in the data processing algorithm: installation/configuration of additional software solutions, downloading/integrating new databases, and editing the overall processing pipeline. In turn, using GRCh37 imposes a number of limitations on the ability to work with actual databases, as their authors focused on a fresh build GRCh38. If we consider the task of variant calling, it is possible to re-run the mapping and variant calling procedure for the up-to-date GRCh38 if raw FASTQ files are available. This process may require additional computational and time resources, which is often not always possible. As an alternative to this solution, we can consider the most resource-efficient procedure, namely genomic coordinate conversion. Considering the above, as well as the fact that NCBI's Remap Tool for genomic coordinate transformation stopped its work in November 2023, the goal was to review existing programs with similar functionality.
Methods and Algorithms: The following bioinformatic tools for genomic coordinates conversion were tested: CrossMap, LiftoverVcf (Picard), transanno, Assembly Converter (Ensembl). To evaluate the performance of the programs, we used 10 VCF files (≈1.5 GB each) that contained an average 4.5 million variants detected through full genome sequencing, as well as the GRCh37_to_GRCh38.chain file required to perform the conversion. Testing was performed in Ubuntu OS (22.04).

Results: The CrossMap was able to convert 98.70 % of all variants in an average of 2 minutes and 17 seconds (±10 seconds) with minimal use of computational resources. The transanno tool showed similar results: 98.72 % of all variants were converted in 2 minutes 7 seconds (±6 seconds). The LiftoverVcf (Picard) converted 98.67 % of the variants in 1 minute 15 seconds (±5 seconds) on average, but the RAM usage reached 12.8 GB at the peak. The web-based tool Assembly Converter (Ensembl) gives the opportunity to run coordinate conversion in a user-friendly graphical interface. It should be noted that the maximum size of a file uploaded to Assembly Converter should not exceed 50 Mb, which imposes limitations on work with full-exome and full-genome sequencing data.
Conclusion: Four tools for genomic coordinates transformation were reviewed. CrossMap and transanno can be called the most universal choices for this task, since these programs are easy to install and do not require large computing power. Assembly Converter is a good option for specialists with little command line experience. LiftoverVcf (Picard) is convenient in cases when you need to perform conversion faster at the expense of additional computational resources.

Список литературы/References

1. Ormond C. et al. Converting single nucleotide variants between genome builds: from cautionary tale to solution. *Briefings Bioinf.* 2021;22(5):bbab069. doi 10.1093/bib/bbab069

doi 10.18699/bgrs2024-12.3-03

Построение однозначной функциональной аннотации генов с помощью метаанализа на основе графов

Бузанов Г.[1, 2]*, Макеев В.[1]

[1] *Институт общей генетики РАН, Москва, Россия*

[2] *Московский физико-технический институт, Москва, Россия*

* *gleb.buzanov@phystech.edu*

Ключевые слова: кластеризация генов; графовые алгоритмы; функциональная аннотация; Enrichr; OpenOrd; функции генов

Мотивация и цель: Исследования GWAS обычно дают ограниченное количество маркерных генов, недостаточно большое для проведения анализа, подобного GSEA. В этой работе мы предлагаем графовый алгоритм, который присваивает каждому гену одну функцию, обеспечивая тем самым однозначную функциональную аннотацию. Для этого большое количество функциональных сигнатур объединяются в единый взвешенный граф, который разбивается на кластеры. Для каждого маркерного гена формируется ограниченный набор генов, которые коэкспрессируются с маркерным геном в большом количестве исследований по профайлингу экспрессии и могут рассматриваться как первичные кандидаты на участие в одних и тех же процессах формирования фенотипа.

Методы и алгоритмы: Основным алгоритмом в работе является алгоритм построения графовой развертки OpenOrd, а также алгоритм K-Means для автоматизации процесса построения кластеризации. Результирующие кластеризации сравнивались по метрикам ассортативности и по мере Жаккара.

Результаты: Полученный алгоритм продемонстрировал лучшее качество кластеризации среди применимых к этой задаче алгоритмов кластеризации, таких как HDBSCAN, k-medoids и др., на выборках размером до 5000 генов. Кроме того, в отличие от других алгоритмов кластеризации, OpenOrd подходит и для более широких выборок, в результате чего возможна кластеризация полного набора из ~20 000 генов, доступных в базах Enrichr. Мы использовали Metascape для аннотирования генов, для выявления доказательств, подтверждающих рациональность кластеризации, и для получения информации о релевантности кластера.

В дальнейшем полученная кластеризация была заново агрегирована алгоритмом для получения иерархической структуры кластеризации. В результате двух дополнительных итераций было получено три слоя кластеризации. На менее детализированных слоях неизбежно возникает проблема слияния слабосвязанных генов в гиперкластеры. Учитывая, что иерархическая многослойная кластеризация логически выполняется на обширных наборах данных с большим количеством генов, можно сделать вывод, что такая организация данных оказывается неэффективной в данной конкретной задаче.

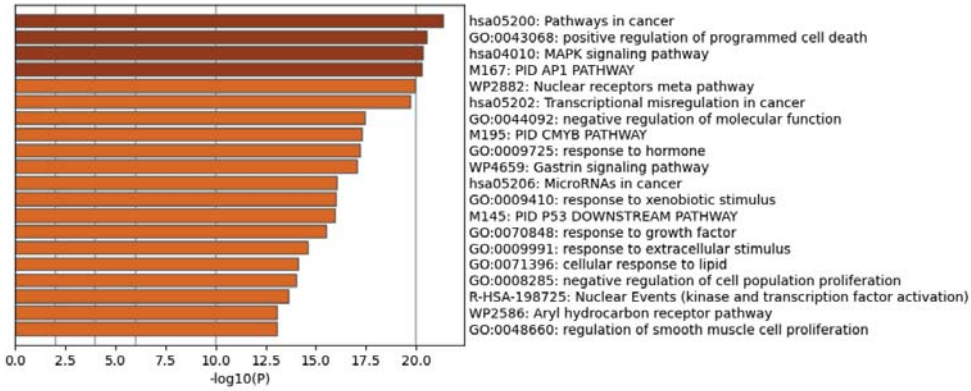


Figure 8: TP53 cluster analysis

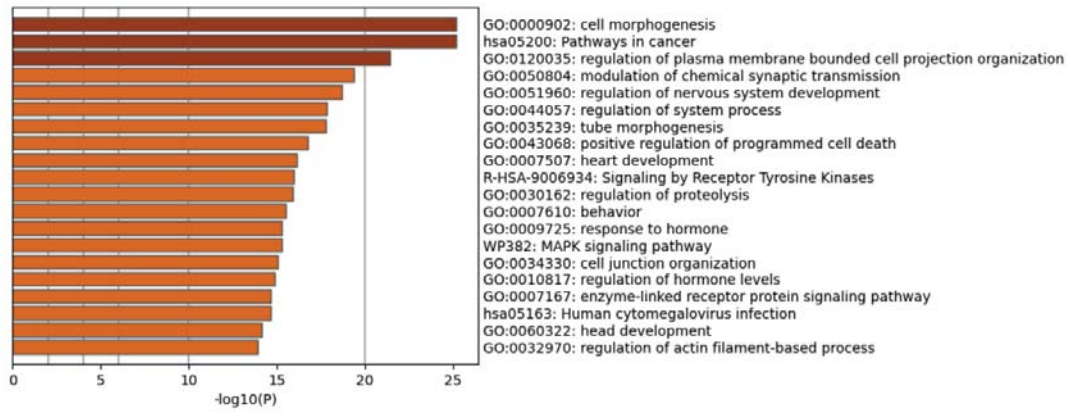


Figure 9: MYH14 cluster analysis

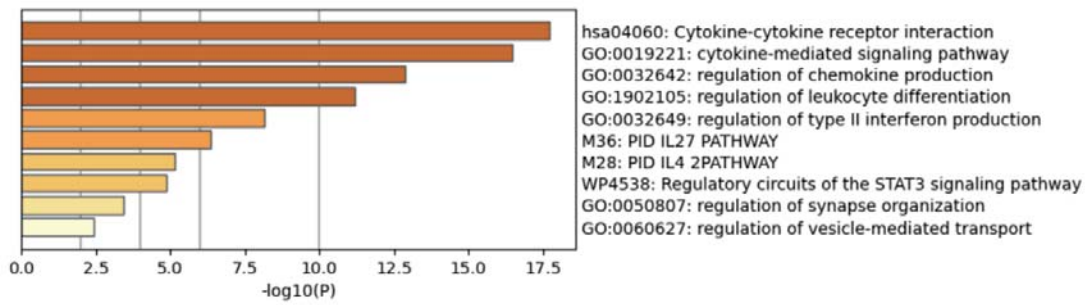


Figure 10: IL6 cluster analysis

Constructing an exclusive gene functional annotation with the help of a graph-based metaanalysis

Buzanov G.[1, 2]*, Makeev V.[1]

[1] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, Moscow, Russia*

* *gleb.buzanov@phystech.edu*

Key words: gene clustering; graph-based algorithm; functional annotation; Enrichr database; OpenOrd; gene functions

Motivation and Aim: GWAS studies usually produce a limited numbers of marker genes, not large enough to provide GSEA-like analysis. Here we suggest a graph based algorithms that assignes a single function to each gene, thus providing an exclusive functional annotation. To this end a large number of functional signatures are combined into a single weighted graph, which is segregated into cliques. For each marker gene a limited gene set is formed that are coexpressed with the marker gene in the large number of co-expression studies and can be considered as a primary candidates for taking part in the same processes of phenotype formation.

Methods and Algorithms: The main algorithm in the work is the OpenOrd graph scan algorithm, as well as the K-Means algorithm for automating the clustering construction process. The resulting clusterings were compared using assortativity metrics and the Jaccard measure.

Results: The resulting algorithm demonstrated the best quality of clustering among clustering algorithms applicable to this problem, such as HDBSCAN, k-medoids, etc. on samples of up to 5,000 genes. In addition, unlike other clustering algorithms, OpenOrd is also suitable for wider samples, as a result of which it is possible to cluster the entire set of about 20,000 genes available in the Enrichr databases. We used Metascape to annotate genes, to identify evidence supporting the rationality of clustering, and to obtain information about cluster relevance.

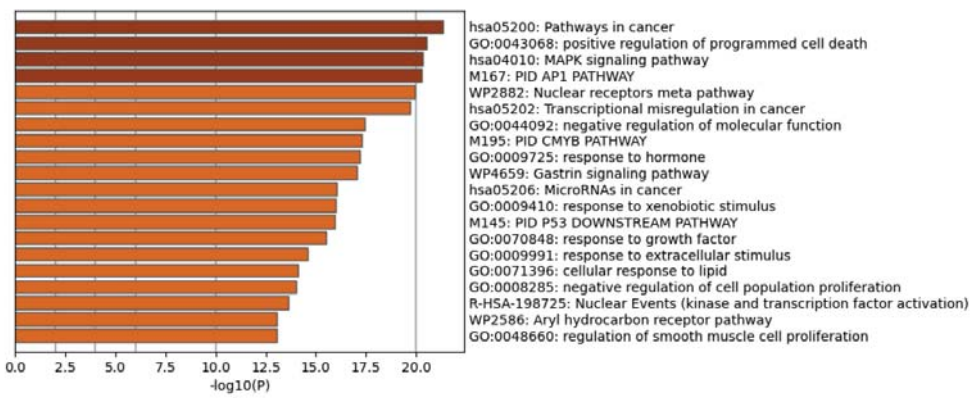


Figure 8: TP53 cluster analysis

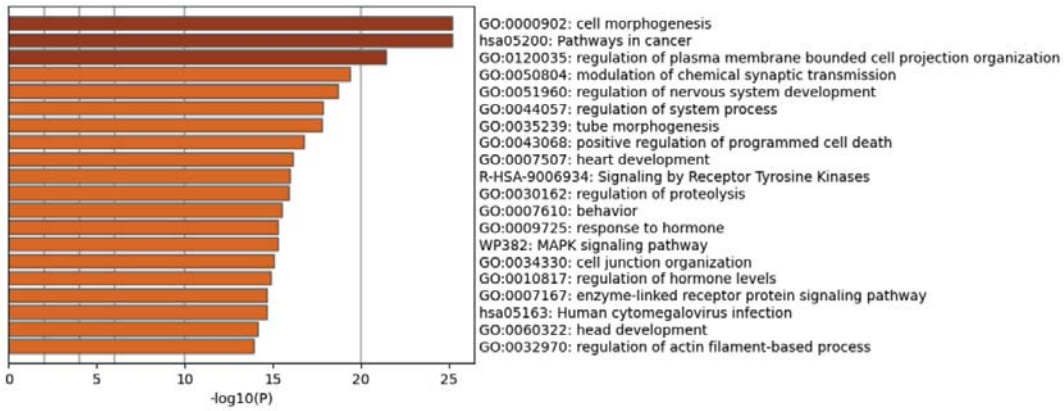


Figure 9: MYH14 cluster analysis

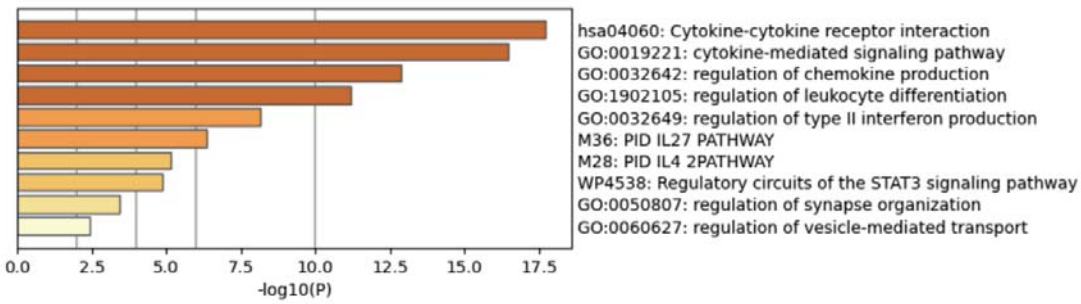


Figure 10: IL6 cluster analysis

The resulting clustering was then re-aggregated by the algorithm to obtain a hierarchical clustering structure. As a result of two additional iterations, three clustering layers were obtained. At less detailed layers, the problem of merging loosely connected genes into hyperclusters inevitably arises. Given that hierarchical multilayer clustering logically works on large datasets with large numbers of genes, it can be concluded that such data organization is ineffective for this particular task.

Список литературы/References

1. Ashburner M. et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25(1):25-29
2. Ashworth A., Lord C.J., Reis-Filho J.S. Genetic interactions in cancer progression and treatment. *Cell*. 2011;145(1):30-38
3. Cano-Gamez E., Trynka G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet*. 2020;11:424
4. Chen E.Y., Tan C.M., Kou Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128
5. Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Res*. 2001;11(8):1425-1433
6. Harris M.A., Clark J., Ireland A. et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32(Database issue):D258-D261

7. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330-D338
8. Gu J., Dai J., Lu H., Zhao H. Comprehensive Analysis of Ubiquitously Expressed Genes in Humans from A Data-driven Perspective. *Genomics Proteomics Bioinformatics*. 2023;21(1):164-176
9. Hochane M., van den Berg P.R., Fan X. et al. Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development. *PLoS Biol*. 2019;17(2):e3000152
10. Kuleshov M.V., Jones M.R., Rouillard A.D. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90-W97
11. Li X., Guo C., Ahmad S. et al. Systematic Analysis of MYB Family Genes in Potato and Their Multiple Roles in Development and Stress Responses. *Biomolecules*. 2019;9(8):317
12. Mar'n S., Brown W.M., Klavans R., Boyack K. OpenOrd: an open-source toolbox for large graph layout. In: Visualization and Data Analysis. 2011;7868:786806
13. Culhane A.C., Schwarzl T., Sultana R. et al. GeneSigDB – a curated database of gene expression signatures. *Nucleic Acids Res*. 2010;38(Database issue):D716-D725
14. Liberzon A., Subramanian A., Pinchback R. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740

doi 10.18699/bgrs2024-12.3-04

Автоматизация детекции центров колосков на RGB изображениях колоса пшеницы

Бусов И.Д.[1, 2]*, Генаев М.А.[1, 3], Комышев Е.Г.[1], Пронина В.А.[2], Коваль В.С.[1, 3], Афонников Д.А.[1, 2, 3]

[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[2] *Новосибирский государственный университет, Новосибирск, Россия*

[3] *Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия*

* *i.busov@g.nsu.ru*

Ключевые слова: пшеница; колос; подсчет колосков; сегментация; детекция; компьютерное зрение; глубокое машинное обучение

Мотивация и цель: Пшеница – это важная сельскохозяйственная культура, которая является одним из основных продуктов питания. Количество колосков в колосе растения является значимым признаком при селекции пшеницы, поскольку это свойство непосредственно связано с урожайностью. Помимо количества колосков, их взаимное расположение также может иметь биологическое значение. Цель работы заключается в разработке точного и эффективного метода для автоматического обнаружения и подсчета центров колосков пшеницы на RGB изображениях колоса.

Методы и алгоритмы: В рамках данного исследования были рассмотрены несколько подходов к решению поставленной задачи: 1) использование нейронной сети с архитектурой YOLO [1], с ограничивающими квадратами с фиксированной длиной стороны, в центре которых находились центры колосков; 2) использование нейронной сети с архитектурой U-Net [2], обученной на бинарных масках с использованием бинарной кросс-энтропии [3]; 3) использование нейронной сети с архитектурой U-Net, обученной на масках из нормального двумерного распределения с применением дивергенции Кульбака–Лейблера [4] в качестве функции потерь. Гиперпараметры моделей подбирались с использованием Байесовской оптимизации (Гауссовских процессов) [5]. Для оценки качества моделей использовались пространственные метрики детекции колоска на изображении (precision, recall, f1) и количественные метрики предсказания числа колосков в колосе (MAE, MSE). Пространственные метрики считались на основе характеристик TP, FP и FN. Случай TP – это предсказанное положение центра колоска, находящееся в радиусе 3 мм от ближайшего реального центра. Если количество предсказанных центров, находящихся в этой окрестности радиуса 3 мм, превышало 1, то оставшиеся предсказанные центры определялись как FP. В случае, если предсказанные центры не находились в указанной окрестности, они относились к FN.

Результаты: Обученные модели были протестированы на двух выборках: тестовой и отложенной. Тестовая выборка состояла из изображений, полученных одновременно с изображениями, на которых обучались нейронные сети, выборка включала в себя 348 изображений колосьев пшеницы. Отложенная выборка была получена намного позже тренировочной и включала в себя 14 изображений колосьев пшеницы. Результаты на тестовой выборке представлены в табл. 1,

результаты на отложенной выборке представлены в табл. 2. Наилучшие результаты продемонстрировала модель с архитектурой U-Net, обученная на нормальных масках с помощью дивергенции Кульбака–Лейблера в качестве функции потерь. Модель с архитектурой U-Net, обученная на бинарных масках с помощью бинарной кросс-энтропии в качестве функции потерь, продемонстрировала результаты немного хуже. Модель, основанная на архитектуре YOLO, показала наихудшие результаты.

Таблица 1. Результаты на тестовой выборке

модель (3 мм)	precision	recall	f1	mae	mse
U-Net (kld)	0.977	0.976	0.975	0.557	1.086
U-Net (bce)	0.983	0.963	0.970	0.597	1.551
YOLO	0.733	0.760	0.734	3.641	26.486

Примечание. Здесь и в табл. 2: U-Net (kld) – модель нейронной сети с архитектурой U-Net, обученной на масках из нормального двумерного распределения с применением дивергенции Кульбака–Лейблера в качестве функции потерь; U-Net (bce) – нейронной сети с архитектурой U-Net, обученной на бинарных масках с использованием бинарной кросс-энтропии; YOLO – нейронной сети с архитектурой YOLO

Таблица 2. Результаты на отложенной выборке

модель (3 мм)	precision	recall	f1	mae	mse
U-Net (kld)	0.971	0.953	0.959	0.929	2.071
U-Net (bce)	0.975	0.946	0.957	1.071	3.214
YOLO	0.678	0.753	0.704	3.0	15.714

Выводы: В ходе работы была достигнута поставленная цель, заключающаяся в разработке точного автоматического метода для локализации и подсчета колосков пшеницы на изображениях.
Финансирование: Исследование поддержано грантом РНФ № 23-14-00150.

Automation of spikelet center detection on RGB images of wheat ear

Busov I.D.[1, 2]*, Genaev M.A.[1, 3], Komyshev E.G.[1], Pronina V.A.[2], Koval V.S.[1, 3], Afonnikov D.A.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *i.busov@g.nsu.ru*

Key words: wheat; spikelet; spikelet counting; segmentation; detection; computer vision; deep machine learning

Motivation and Aim: Wheat is an important crop that is a staple food. The number of spikelets in an ear of a plant is a significant trait in wheat breeding because this trait is

directly related to yield. In addition to the number of spikelets, their mutual arrangement may also have biological significance. The aim of this work is to develop an accurate and efficient method for automatic detection and counting of wheat spikelet centers in RGB images of the spikelet.

Methods and Algorithms: In this research, several approaches have been considered to solve the problem: (1) using a neural network with the YOLO architecture [1], with bounding squares with fixed side lengths at the center of which were the spikelet centers; (2) using a neural network with the U-Net architecture [2] trained on binary masks using binary cross-entropy [3]; and (3) a neural network with U-Net architecture trained on masks from a normal bivariate distribution using the Kulback–Leibler divergence [4] as a loss function. The hyperparameters of the models were selected using Bayesian optimization (Gaussian processes) [5]. Spatial metrics for spikelet detection in the image (precision, recall, f1) and quantitative metrics for predicting the number of spikelets in a spikelet (MAE, MSE) were used to evaluate the quality of the models. The spatial metrics were considered based on TP, FP and FN features. The TP case is the predicted position of the spikelet center that is within 3 mm radius from the nearest real center. If the number of predicted centers located in this vicinity of 3 mm radius was greater than 1, the remaining predicted centers were defined as FP. In case the predicted centers were not in the specified vicinity, they were referred to as FN.

Results: The trained models were tested on two samples: a test sample and a delayed sample. The test sample consisted of images acquired simultaneously with the imagesb on which the neural networks were trained, the sample included 348 images of wheat ears. The delayed sample was obtained much later than the training sample, and included 14 images of wheat ears. The results on the test sample are presented in Table 1, and the results on the deferred sample are presented in Table 2. The model with U-Net architecture trained on normal masks using Kulbak–Leibler divergence as the loss function showed the best results. The model with U-Net architecture trained on binary masks using binary cross-entropy as a loss function performed slightly worse. The model based on YOLO architecture showed the worst results.

Table 1. Results on the test sample

model (3 мм)	precision	recall	f1	mae	mse
U-Net (kld)	0.977	0.976	0.975	0.557	1.086
U-Net (bce)	0.983	0.963	0.970	0.597	1.551
YOLO	0.733	0.760	0.734	3.641	26.486

Note. Here and in Table 2: U-Net (kld) – a neural network with U-Net architecture trained on masks from a normal bivariate distribution using Kullback–Leibler divergence as the loss function. U-Net (bce) – neural network with U-Net architecture trained on binary masks using binary cross-entropy. YOLO – neural network with YOLO architecture

Table 2. Results on delayed sampling

model (3 мм)	precision	recall	f1	mae	mse
U-Net (kld)	0.971	0.953	0.959	0.929	2.071
U-Net (bce)	0.975	0.946	0.957	1.071	3.214
YOLO	0.678	0.753	0.704	3.0	15.714

Conclusions: The work achieved its objective of developing an accurate automatic method for localizing and counting wheat spikelets in images.
Funding: The research was supported by the RSF grant No. 23-14-00150.

Список литературы/References

1. Jiang P., Ergu D., Liu F., Cai Y., Ma B. A Review of Yolo algorithm developments. *Procedia Comput Sci.* 2022;199:1066-1073
2. Siddique N., Paheding S., Elkin C.P., Devabhaktuni V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access.* 2021;9:82031-82057
3. Ruby U., Yendapalli V. Binary cross entropy with deep learning technique for image classification. *Int J Adv Trends Comput Sci Eng.* 2020;9(4):5393-5397
4. Kim T., Oh J., Kim N., Cho S., Yun S.Y. Comparing kullback-leibler divergence and mean squared error loss in knowledge distillation. *arXiv.* 2021
5. MacKay D.J. Introduction to Gaussian processes. 1998. [http://www.inference.org.uk/mackay/gpB.pdf]

doi 10.18699/bgrs2024-12.3-05

Реализация в задачном подходе к Искусственному Интеллекту идеи П.К. Анохина об обобщении Теории Функциональных Систем на другие области знаний

Витяев Е.Е.
Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия
vityaev@math.nsc.ru

Ключевые слова: теория функциональных систем; задача; искусственный интеллект

Мотивация и цель: В настоящее время широкое развитие получил ***Общий искусственный интеллект (Artificial General Intelligence – AGI)***, основанный на том, что моделирование когнитивных процессов человека не является обязательным условием решения интеллектуальных задач в силу результатов глубокого обучения и больших лингвистических моделей. Соответственно, возникает понятие «Общего Искусственного Интеллекта», которым потенциально может обладать, как человек или живой организм с высокоразвитой центральной нервной системой, так и абстрактная робототехническая система. В [1] дается следующее определение AGI: «Это способность решать когнитивные задачи в целом, действуя целенаправленно, адаптируясь к условиям среды через обучение, минимизируя риски и оптимизируя потери на достижение своих целей».
Цель доклада показать, что предлагаемый нами Задачный подход к Искусственному Интеллекту не только решает сформулированную перед AGI задачу, но, что принципиально более важно, моделирует когнитивные процессы человека и поэтому является вариантом «Естественного интеллекта». Как показано в докладе он достаточно точно моделирует целенаправленную деятельность человека в соответствии с Теорией Функциональных Систем (ТФС) работы мозга [2]. Тем самым задачный подход в достаточно полной мере реализует мечту П.К. Анохина об обобщении ТФС на другие области знаний.
Методы и алгоритмы: ***Задачный подход. Понятие «задача» в основаниях математики*** [3]**.** Анализ понятия задачи начинается со следующих рассуждений д.ф.н. К.Ф. Самохвалова: «Я хочу пить» – что это значит? Нет, конечно, никакой ошибки полагать, что слова «я хочу пить» означают определенное состояние сознания, именуемое жаждой. Но тогда возникает новый вопрос: как ощущение жажды (хотения) связано с фактическим питьем (удовлетворением хотения)? Откуда я знаю, что удовлетворить жажду можно питьем? Содержится ли в самом переживании жажды сознание того, чем эту жажду можно удовлетворить? … *Знать желание не означает знать желаемое, а означает способность узнать желаемое*», т. е. знать и иметь критерий удовлетворения желания.
Таким образом, задача определена (осмыслена) тогда и только тогда, когда есть критерий решенности задачи – критерий проверки действительно ли предъявленное решение является решением задачи. В [3] было доказано, что только в «слабых» формальных системах (для которых не проходит теорема Гёделя) мы можем средствами самой формальной системы определить, является ли некоторый текст доказательством решения задачи или нет. В результате

программа Гильберта обоснования математики формулируется иначе: не нужно для всей математики доказывать её непротиворечивость – это невозможно и ненужно. Надо решать задачи в рамках слабых формальных систем.

Понятие «задача» в когнитивных науках [4]. Обобщением понятия задача в когнитивных науках является понятие ***Цели***. Цель нельзя достичь, не имея критерия её достижения, иначе всегда можно считать, что Цель уже достигнута. Поэтому формулировка Цели всегда должна включать некоторый критерий достижения цели. Достижение Цели дает Результат. Единственной физиологической теорией, в которой достижение Цели и получение Результата рассматривается как *решение мозгом Задачи* по удовлетворению некоторой потребности, является Теория Функциональных Систем [2]. П.К. Анохин писал: «Пожалуй, одним из самых драматических моментов в истории изучения мозга как интегративного образования является фиксация внимания на самом действии, а не на его результатах … мы можем считать, что результатом «хватательного рефлекса» будет не само хватание как действие, а та совокупность афферентных раздражений, которая соответствует признакам «схваченного» предмета». «Совокупность афферентных раздражений» и есть критерий достижения цели в ТФС. Целью в ТФС является удовлетворение некоторой потребности. Согласно ТФС центральные механизмы функциональных систем, обеспечивающих целенаправленное поведение, имеют следующую однотипную архитектуру.

Афферентный синтез. Начальную стадию поведенческого акта любой степени сложности составляет афферентный синтез, включающий синтез *мотивационного возбуждения, памяти, обстановочной и пусковой афферентации*.

Мотивационное возбуждение. Постановка цели осуществляется возникшей потребностью, которая трансформируется в мотивационное возбуждение.

*Память***.** Мотивационное возбуждение «извлекает из памяти» все возможные способы достижения цели, а также всю последовательность и иерархию результатов, которые должны быть получены для достижения цели.

Обстановочная и пусковая афферентации фиксируют свойства обстановки и время начала поведенческого акта.

Принятие решений. На стадии афферентного синтеза мотивационным возбуждением может быть извлечено из памяти (в данной обстановке) несколько способов достижения цели. На стадии принятия решения выбирается только один из них. «Вытягивая» из памяти весь накопленный опыт, мотивационное возбуждение преобразуется в *конкретную цель*, определяющую способ своего достижения. Конкретная цель называется в ТФС «*высшей мотивацией*».

Акцептором результатов действия. Мотивационное возбуждение «извлекает из памяти» также всю последовательность и иерархию подрезультатов, которые должны быть получены в процессе выполнения плана действий и достижения цели. Эта последовательность называется в ТФС *акцептором результатов действия*, представляющим собой доминирующую потребность (Цель) организма, трансформированную в форме опережающего возбуждения мозга, как бы в своеобразный комплексный рецептор будущего подкрепления, который является *критерием достижения конкретной цели*.

Подкрепление. **Санкционирующая стадия**. Если в результате выполнения конкретного плана действий цель будет достигнута (потребность удовлетворена) и все подрезультаты акцептора результатов действия также достигнуты, то

возникает последняя санкционирующая стадия, в которой осуществляется удовлетворение потребности и занесение конкретного плана действий в память.

Результаты: **Понятие «задача» в семантическом моделировании.** Распространим, как хотел П.К. Анохин, когнитивный процесс достижения цели в ТФС на процесс решения задач. Для этого уточним понятие «задача». Будем считать, что задача определена тогда и только тогда, когда в ней присутствуют:

– предметная область, зафиксированная в виде модели в некоторой онтологии, а также знания о предметной области, исходные данные, факты и гипотезы;
– запрос (вопрос), сформулированный в задаче, относящийся к предметной области, на который мы должны получить ответ;
– критерий удовлетворения запроса – в каком случае можно считать, что ответ на запрос (вопрос) получен.
– в каком контексте следует искать ответ на запрос (вопрос) – какую цель мы преследуем, решая задачу, т. е. что мы ожидаем от полученного результата.

Решение задачи в ТФС может быть сформулировано так:

– предметная область – внешний мири модель внешнего мира в виде «Образа мира», в котором осуществляется целенаправленное поведение;
– запрос к предметной области – это потребность-цель, которую надо удовлетворить, представленная некоторым целевым предикатом P_0;
– под решением задачи понимается объект, ситуация, событие внешнего мира, воспринимаемые «совокупностью афферентных раздражений» *a*, которое делает предикат P_0 истинным (удовлетворяет потребность). Истинность $P_0(a)$ и есть критерий решения задачи.
– принятие решений. конкретных планов действий по достижению цели P_0 может быть несколько. Во время принятия решений выбирается один план действий с учетом контекста, зафиксированного в модели.

Задачный подход предполагает, что истинным назначением ИИ является автоматизация решения задач и запросы должны формулироваться в терминах исполнимых спецификаций. В качестве концепции базовой модели вычислений предлагается взять концепцию $\sum$-определимости вычислений [5], дополнив ее процедурой проверки истинности $\sum$-формул на конструктивной модели M, рассматриваемой совместно с ее списочной надстройкой HW(M) [6, 7].

Определим теперь в рамках данной концепции понятие «задача». Под решением задачи будем понимать набор констант, делающий $\sum$-формулу запроса при означивании ее переменных константами истинной на модели предметной области. Важно подчеркнуть, что истинность $\sum$- формулы запроса, получаемой подстановкой констант вместо переменных, и есть критерий решенности задачи. Так понимаемый подход к формулировке и решению задач носит название *семантического моделирования*.

Следует отметить, что с практической точки зрения для спецификации задач вполне достаточно использовать не весь $\sum$-язык, а его Δ_0-фрагмент, куда относятся $\sum$-формулы, содержащие в свой записи только ограниченные кванторы всеобщности и существования. Этот Δ_0-фрагмент обладает тем достоинством, что при определенных условиях можно гарантировать полиномиальную вычислимость специфицируемых в этом языке предикатов и функций [8, 9].

Финансирование: Работа выполнена по Государственному Заданию: «Логические исчисления и семантики, теории моделей и вычислимости» FWNF-2022-0011.

The implementation in the Task-driven approach to Artificial Intelligence the P.K. Anokhin's idea of generalizing the Theory of Functional Systems to other fields of knowledge

Vityaev E.
S.L. Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia
vityaev@math.nsc.ru

Key words: theory of functional systems; task; artificial intelligence

Motivation and Aim: Currently, ***Artificial General Intelligence*** (AGI) has been widely developed, based on the fact that modeling human cognitive processes is not a prerequisite for solving intellectual problems due to the results of deep learning and large linguistic models. Accordingly, the concept of "General Artificial Intelligence" arises, which can potentially be possessed by both a person or a living organism with a highly developed central nervous system, and an abstract robotic system. In [1], the following definition of AGI is given: "It is the ability to solve cognitive tasks in general, acting purposefully, adapting to environmental conditions through learning, minimizing risks and optimizing losses to achieve one's goals".

The purpose of the report is to show that our proposed Task Approach to Artificial Intelligence not only solves the problem formulated for AGI, but, more importantly, models human cognitive processes and therefore is a variant of "Natural Intelligence". As shown in the report, he accurately models purposeful human activity in accordance with the Theory of Functional Systems (TFS) of the brain [2]. Thus, the task approach fully realizes P.K. Anokhin's dream of generalizing TFS to other fields of knowledge.

Methods and Algorithms: ***A task-driven approach***. The concept of "task" in the foundations of mathematics [3]. The analysis of the concept of a task begins with the following arguments of Dr. F.N. K.F. Samokhvalov: "I want to drink" – what does this mean? There is, of course, no mistake in assuming that the words "I am thirsty" mean a certain state of consciousness called thirst. But then a new question arises: how is the feeling of thirst (desire) related to the actual drinking (satisfaction of desire)? How do I know that drinking can satisfy thirst? Does the experience of thirst itself contain a consciousness of how this thirst can be satisfied? ... To know desire does not mean to know what is desired, but means the ability to know what is desired", i. e. to know and have a criterion for satisfying desire.

Thus, the problem is defined (understood) if and only if there is a criterion for solving the problem – a criterion for checking whether the presented solution is really a solution to the problem. In [3] it was proved that only in "weak" formal systems (for which Godel's theorem does not pass) we can determine by means of the formal system itself whether a certain text is a proof of the solution of the problem or not. As a result, Hilbert's program for the justification of mathematics is formulated differently: it is not necessary for all mathematics to prove its consistency – this is impossible and unnecessary. It is necessary to solve problems within the framework of weak formal systems.

The concept of "task" in cognitive sciences [4]. A generalization of the concept of a task in cognitive sciences is the concept of a ***Goal***. The goal cannot be achieved without having a criterion for achieving it, otherwise it can always be assumed that the Goal has already been achieved. Therefore, the formulation of the Goal should always include

some criterion for achieving the goal. Achieving a Goal gives a Result. The only physiological theory in which achieving a Goal and obtaining a Result is considered as the brain's solution to the Problem of satisfying some need is the Theory of Functional Systems [2]. P.K. Anokhin wrote: “Perhaps one of the most dramatic moments in the history of studying the brain as an integrative education is the fixation of attention on the action itself, and not on its results... we can assume that the result of the “grasping reflex” will not be grasping itself as an action, but that set of afferent stimuli that corresponds to the signs of the “grasped” object”. The “totality of afferent stimuli” is the criterion for achieving the goal in TFS. The goal in TFS is to satisfy some need. According to the TFS, the central mechanisms of functional systems that ensure purposeful behavior have the following architecture of the same type.

Afferent synthesis. The initial stage of a behavioral act of any degree of complexity is afferent synthesis, including the synthesis of motivational excitement, memory, situational and trigger afferentation.

Motivational excitement. The goal setting is carried out by the ***need*** that has arisen, which is transformed into motivational excitement.

Memory. Motivational excitement “extracts from memory” all possible ways to achieve a goal, as well as the entire sequence and hierarchy of results that must be obtained to achieve the goal.

Situational and trigger afferentation fix the properties of the situation and the time of the beginning of the behavioral act.

Decision-making. At the stage of afferent synthesis, motivational excitement can be extracted from memory (in a given setting) several ways to achieve a goal. At the decision-making stage, only one of them is selected. By “pulling” all the accumulated experience from memory, motivational excitement is transformed into a ***specific goal*** that determines the way to achieve it. A specific goal is called a “higher motivation”.

The acceptor of the actions result. Motivational excitement also “extracts from memory” the entire sequence and hierarchy of sub-goals that must be obtained in the process of implementing the action plan and achieving the goal. This sequence is called in the TFS an acceptor of the results of an action, which represents the dominant need (Goal) of the body, transformed in the form of anticipatory excitation of the brain, as if into a kind of complex receptor for future reinforcement, which is a criterion for achieving a specific goal.

Reinforcements. The sanctioning stage. If, as a result of the implementation of a specific action plan, the goal is achieved (the need is satisfied) and all the sub-results of the acceptor of the action results are also achieved, then the last authorized stage arises, in which the need is satisfied and the specific action plan is stored in memory.

Results: ***The concept of "task" in semantic modeling***. Let’s extend, as P.K. Anokhin wanted, the cognitive process of the goals achievement in TFS to the process of solving problems. To do this, we will clarify the concept of “task”. We will assume that a task is defined if and only if it contains:

– the subject area fixed as a model in some ontology, as well as knowledge about the subject area, initial data, facts and hypotheses;
– a query (question) formulated in the task, related to the subject area, to which we must receive an answer;
– the criterion for satisfying the request is in which case it can be considered that the answer to the request (question) has been received.

– in what context should we look for an answer to the query (question) – what goal are we pursuing when solving the problem, i. e. what do we expect from the result.

The solution of the problem in the TFS can be formulated as follows:

– the subject area is the external world and the model of the external world in the form of an "Image of the world" in which purposeful behavior is carried out;
– a request to the subject area is a need–a goal that needs to be satisfied, represented by some target predicate P_0;
– the solution of the problem is understood as an object, situation, event of the external world, perceived by a "set of afferent stimuli" a, which makes the predicate P_0 true (satisfies the need). The truth of $P_0(a)$ is the criterion for solving the problem.
– decision-making. There may be several specific action plans to achieve the P_0 goal. During decision-making, one action plan is selected, taking into account the context fixed in the model.

The task approach assumes that the true purpose of AI is to automate problem solving and queries should be formulated in terms of executable specifications. As a concept of the basic model of calculations, it is proposed to take the concept of $\sum$-definability of calculations [5], supplementing it with a procedure for verifying the truth of $\sum$-formulas on the constructive model M together with its list superstructure HW(M) [6, 7].

Let us now define the concept of "task" within the framework of this concept. By solving the problem, we will understand a set of constants that makes the $\sum$-formula of the query when denoting its variables by constants true on the domain model. It is important to emphasize that the truth of the query formula obtained by substituting constants instead of variables is the criterion for solving the problem. The so-understood approach to problem formulation and solution is called semantic modeling.

It should be noted that from a practical point of view, for the specification of tasks, it is quite sufficient to use not the entire $\sum$-language, but its Δ_0-fragment, which includes $\sum$-formulas containing only limited quantifiers of universality and existence. This Δ_0 fragment has the advantage that, under certain conditions, it is possible to guarantee the polynomial computability of predicates and functions specified in this language [8, 9].

Funding: The work was performed according to the State Assignment: "Logical calculus and semantics, model theory and computability" FWNF-2022-0011.

Список литературы/References

1. Ведяхин А. и др. Сильный искусственный интеллект. На подступах к сверхразуму. М.: Интеллектуальная Литература, 2021
2. Анохин П.К. Принципиальные вопросы теории функциональных систем. В: Философские аспекты теории функциональных систем. М.: Наука, 1978;49-106
3. Ершов Ю.Л., Самохвалов К.Ф. Современная философия математики. Новосибирск, 2007
4. Vityaev E.E. Purposefulness as a Principle of Brain Activity. In: Nadin M. (Ed.). Anticipation: Learning from the Past. Cognitive Systems Monographs. V. 25. Springer, 2015;231-254
5. Ершов Ю.Л. Определимость и вычислимость. Новосибирск, 2000
6. Гончаров С.С., Свириденко Д.И. Семантическое моделирование и искусственный интеллект. *Сибирский философский журнал*. 2018;16(4):5-25
7. Goncharov S.S. Conditional terms in semantic programming. *Sib Math J*. 2017;58(5):794-800
8. Гончаров С.С., Свириденко Д.И. Логический язык описания полиномиальной вычислимости. *Доклады РАН*. 2018;485(1):11-14
9. Goncharov S., Ospichev S., Ponomaryov D., Sviridenko D. The expressiveness of looping terms in the Semantic Programming. *Сибирские электронные математические известия.* 2020;17:380-394

doi 10.18699/bgrs2024-12.3-06

Поиск ДНК-связывающих белков (DBPs) с использованием методов глубокого обучения

Гавриленко А.[1]*, Шабурова Е.[1], Антонец Д.[1], Вяткин Ю.[1], Раменский В.[1, 2, 3]

[1] *Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем, Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

[2] *Национальный медицинский исследовательский центр терапии и профилактической медицины Минздрава России, Москва, Россия*

[3] *Факультет биоинженерии и биоинформатики, Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

* *a.gavrilenko@iai.msu.ru*

Ключевые слова: глубокое обучение; протеомика; языковые модели для белков; Ankh; ДНК-связывающие белки

Мотивация и цель: ДНК-связывающие белки (DBPs) играют ключевую роль во множестве биологических процессов, таких как репликация, репарация ДНК, регуляция транскрипции и трансляция [1]. Приблизительно 6–7 % генов в геномах эукариот и 2–5 % в геномах прокариот кодируют ДНК-связывающие белки [2]. Для изучения взаимодействия белков и ДНК используют несколько экспериментальных методов – поверхностный плазмонный резонанс, фильтрационный связывающий анализ, электрофоретический анализ сдвига подвижности [3]. Однако эти подходы имеют свои ограничения, являются длительными и трудоемкими, результаты экспериментов могут быть искажены. Достижения в области высокопроизводительных методов секвенирования значительно увеличили количество геномных данных. Однако менее 1 % из более чем 200 миллионов белков в UniProtKB имеют экспериментально подтвержденные аннотации. Это подчеркивает важность разработки вычислительных методов для автоматического поиска ДНК-связывающих белков (DBP). В последнее время в области вычислительной биологии получили распространение предобученные языковые модели для белков, многие из которых аналогичны по архитектуре языковым моделям для обработки естественного языка [4]. Они способны кодировать так называемые представления белка по его аминокислотной последовательности в виде вектора чисел, отражающего его свойства и функции. Эти представления используются для решения различных задач: классификация вторичной [5], третичной структуры белка [6], предсказания функции ферментов [7].

В настоящее время не существует общепринятых методов для предсказания способности белков связываться с ДНК. По этой причине использование предобученных белковых языковых моделей для решения данной задачи представляет значительный научный интерес. Целью нашей работы была разработка алгоритма для поиска белков, способных связываться с ДНК.

Методы и алгоритмы: Для формирования обучающего набора данных использовались аминокислотные последовательности белков из базы данных UniProtKB/Swiss-Prot [8]. Каждая запись в базе данных вручную проверяется и аннотируется специалистами, что обеспечивает высокую точность данных.

Задача поиска DBPs формулируется как задача бинарной классификации, где положительный класс составляют ДНК-связывающие белки, отрицательный класс – ДНК-несвязывающие белки.

Белки положительного класса выбирались из Swiss-Prot по имеющейся аннотации Gene Ontology “DNA binding”. Белки отрицательного класса также выбирались из Swiss-Prot, не имеющие аннотацию Gene Ontology “DNA-binding/RNA-binding”. РНК-связывающие белки не включались в отрицательный класс, поскольку многие из них также связываются с ДНК [9]. Для обеспечения уникальности набора данных были удалены идентичные последовательности белков. Белки фильтровались по длине аминокислотных последовательностей: исключались белки, имеющие длину менее 50 или более 1024 аминокислот. Этот подход был применен для удаления потенциальных пептидов и белковых комплексов, а также в связи с вычислительными ограничениями языковых моделей, которые не могут работать с длинными последовательностями белков [10]. Кроме того, исключались белки, имеющие неопределенные аминокислоты “X”. Тестовый набор данных PDB2272 был взят из предыдущего исследования [11], он представляют собой аминокислотные последовательности белков из UniprotKB с имеющейся аннотацией Gene Ontology. Данные использовались для оценки точности классификации разработанного алгоритма.

Алгоритм состоит из двух компонент:

1. Предобученная языковая модель для белков Ankh [12] для получения представлений белков по их аминокислотной последовательности.
2. Классификатор на основе градиентного бустинга, построенный с использованием LightAutoML [13], вычисляющий вероятность связывания ДНК белком.

Результаты: Для предотвращения гомологии между обучающим набором и тестовым набором данных была выполнена кластеризация с использованием MMseqs2 [14] с порогом идентичности аминокислотных последовательностей более чем 50 %. В результате белки из обучающего набора данных, попавшие в один кластер с белками из тестового набора, исключались. Для балансировки обучающего набора данных случайным образом выбирались белки отрицательного класса в эквивалентном соотношении к количеству белков положительного класса. Обучающий набор данных в итоге включил 31 803 белка положительного класса и 31 803 белка отрицательного класса. Информация о обучающем и тестовом наборах данных представлена в табл. 1.

Таблица 1. Статистика обучающего и тестового наборов данных

Набор данных	Количество белков			Количество кластеров
	общее	положительного класса	отрицательного класса	
Обучающий набор данных	63 606	31 803	31 803	28 033
PDB2272	2272	1119	1153	2270

Для оценки точности классификации использовались метрики: Precision, Sensitivity, Specificity, Matthews correlation coefficient (MCC). В ходе исследования

точность классификации алгоритма сравнивалась на тестовом наборе данных PDB2272 с существующими методами (табл. 2).

Таблица 2. Сравнение разработанного нами метода с существующими на тестовом наборе данных (PDB2272)

Алгоритм	Precision	Sensitivity	Specificity	MCC
PHMMER [15]	0.608	0.805	0.466	0.289
BiCaps-DBP [16]	*	0.893	0.707	0.613
Local-DPP [17]	*	0.087	0.936	0.456
Наш алгоритм	0.816	0.859	0.8	0.661

*Авторы работ не предоставили значения соответствующих метрик.

Выводы: В этом исследовании представлен новый метод использования предобученных языковых моделей для белков для идентификации ДНК-связывающих белков, расширяющий применение технологий машинного обучения в молекулярной биологии. Алгоритм показывает высокую эффективность по сравнению с существующими методами, значительно ускоряя и удешевляя процесс идентификации ДНК-связывающих белков. Это способствует более быстрому и точному пониманию их функций, с потенциальным применением в медицинских и биотехнологических исследованиях.

Searching for DNA-binding proteins (DBPs) using deep learning methods

Gavrilenko A.[1]*, Shaburova E.[1], Antonets D.[1], Vyatkin Y.[1], Ramensky V.[1, 2, 3]

[1] *Institute for Artificial Intelligence, Lomonosov Moscow State University, Moscow, Russia*

[2] *National Medical Research Center for Therapy and Preventive Medicine of the Ministry of Healthcare of Russian Federation, Moscow, Russia*

[3] *Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

** a.gavrilenko@iai.msu.ru*

Key words: deep learning; proteomics; language models for proteins; Ankh; DNA-binding proteins

Motivation and Aim: DNA-binding proteins (DBPs) play a crucial role in numerous biological processes such as DNA replication, repair, transcriptional regulation, and translation [1]. Approximately 6–7 % of genes in eukaryotic genomes and 2–5 % in prokaryotic genomes encode DNA-binding proteins (DBPs) [2]. Several experimental methods are employed to study protein-DNA interactions, including surface plasmon resonance, filter binding assays, and electrophoretic mobility shift assays [3]. However, these approaches have limitations, are time-consuming and labor-intensive, and experimental results may be biased. Advances in high-throughput sequencing methods have significantly increased the amount of genomic data available. Nevertheless, less

than 1 % of the over 200 million proteins in UniProtKB have experimentally validated annotations. This underscores the importance of developing computational methods for the automatic identification of DNA-binding proteins (DBPs). Recently, pretrained protein language models have become popular in computational biology, many of which are analogous in architecture to natural language processing models [4]. These models can encode protein representations from their amino acid sequences into vector forms reflecting their properties and functions. These representations are used for various tasks, such as the classification of secondary [5] and tertiary protein structures [6], and the prediction of enzyme functions [7].

Currently, no standardized methods exist for predicting the DNA-binding capability of proteins. For this reason, utilizing pretrained protein language models for this task presents significant scientific interest. The aim of this study was to develop an algorithm for identifying proteins capable of binding DNA.

Methods and Algorithms: Amino acid sequences of proteins from the UniProtKB/Swiss-Prot database [8] were used to form the training dataset. Each entry in the database is manually curated and annotated by experts, ensuring high data accuracy. The task of identifying DBPs is formulated as a binary classification problem, where the positive class comprises DNA-binding proteins and the negative class comprises non-DNA-binding proteins.

Proteins in the positive class were selected from Swiss-Prot based on the existing Gene Ontology annotation "DNA binding." Proteins in the negative class were also selected from Swiss-Prot, excluding those with the Gene Ontology annotation "DNA-binding/RNA-binding." RNA-binding proteins were not included in the negative class, as many of them also bind to DNA [9]. To ensure dataset uniqueness, identical protein sequences were removed. Proteins were filtered by the length of their amino acid sequences, excluding those shorter than 50 or longer than 1,024 amino acids. This approach was applied to remove potential peptides and protein complexes and due to computational limitations of language models, which cannot process long protein sequences [10]. Additionally, proteins with undefined amino acids "X" were excluded. The test dataset PDB2272 was taken from a previous study [11] and consists of amino acid sequences of proteins from UniProtKB with existing Gene Ontology annotations. The data were used to assess the classification accuracy of the developed algorithm.

The algorithm comprises two components:

1. A pretrained protein language model, Ankh [12], to obtain protein representations from their amino acid sequences.
2. A gradient boosting-based classifier built using LightAutoML [13], which calculates the probability of a protein binding to DNA.

The algorithm consists of two components:

1. A pre-trained protein language model, Ankh [12], to extract features from the amino acid sequences.
2. A gradient boosting classifier, built using LightAutoML [13], to predict the probability of a protein being DNA-binding.

Results: To avoid homology between the training and test datasets, clustering was performed using MMseqs2 [14] with an amino acid sequence identity threshold of over 50 %. As a result, proteins from the training dataset that clustered with proteins from the test dataset were excluded. To balance the training dataset, proteins from the negative class were randomly selected in a ratio equivalent to the number of proteins in the positive class. The final training dataset included 31,803 proteins in the positive class

and 31,803 proteins in the negative class. Information about the training and test datasets is presented in Table 1.

Table 1. Statistics of the training and test datasets

Dataset	Number of Proteins	Number of Positive Class Proteins	Number of Negative Class Proteins	Number of Clusters
Training dataset	63,606	31,803	31,803	28,033
PDB2272	2,272	1,119	1,153	2,270

Metrics such as Precision, Sensitivity, Specificity, and Matthews correlation coefficient (MCC) were used to evaluate classification accuracy. During the study, the classification accuracy of the algorithm was compared with existing methods on the PDB2272 test dataset (Table 2).

Table 2. Comparison with existing methods on an independent test set (PDB2272)

Algorithm	Precision	Sensitivity	Specificity	MCC
PHMMER [15]	0.608	0.805	0.466	0.289
BiCaps-DBP [16]	*	0.893	0.707	0.613
Local-DPP [17]	*	0.087	0.936	0.045
Our Algorithm	0.816	0.859	0.800	0.661

*The authors did not provide values for corresponding metrics.

Conclusion: This study introduces a novel method using pretrained protein language models to identify DNA-binding proteins, extending machine learning applications in molecular biology. The algorithm shows high efficiency compared to existing methods, significantly accelerating and reducing the cost of identifying DNA-binding proteins. This contributes to a faster and more accurate understanding of their functions, with potential applications in medical and biotechnological research.

Список литературы/References

1. Ahmed S., Kabir M., Ali Z. et al. An Integrated Feature Selection Algorithm for Cancer Classification using Gene Expression Data. *Comb Chem High Throughput Screen.* 2018;21(9):631-645
2. Luscombe N.M., Austin S.E., Berman H.M., Thornton J.M. An overview of the structures of protein-DNA complexes. *Genome Biol*. 2000;1(1):REVIEWS001
3. Ferraz R.A.C., Lopes A.L.G., da Silva J.A.F. et al. DNA-protein interaction studies: A historical and comparative analysis. *Plant Methods.* 2021;17(1):82
4. Ofer D., Brandes N., Linial M. The language of proteins: NLP, machine learning & protein sequences. *Comput Struct Biotechnol J.* 2021;19:1750-1758

5. Rives A., Meier J., Sercu T. et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc Natl Acad Sci USA.* 2021;118(15):e2016239118
6. Chowdhury R, Bouatta N, Biswas S. et al. Single-sequence protein structure prediction using a language model and deep learning. *Nat Biotechnol.* 2022;40(11):1617-1623
7. Yu T., Cui H., Li J.C. et al. Enzyme function prediction using contrastive learning. *Science.* 2023;379(6639):1358-1363
8. UniProt Consortium T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2018;46(5):2699
9. Hudson W.H., Ortlund E.A. The structure, function and evolution of proteins that bind DNA and RNA. *Nat Rev Mol Cell Biol.* 2014;15(11):749-760
10. Tan N.Ö., Peng A.Y., Bensemann J. et al. Input-length-shortening and text generation via attention values. *arXiv*. 2023
11. Lou W., Wang X., Chen F. et al. Sequence based prediction of dna-binding proteins based on hybrid feature selection using random forest and gaussian naïve bayes. *PLoS One.* 2014;9(1):e86703
12. Elnaggar A., Essam H., Salah-Eldin W. et al. Ankh: Optimized protein language model unlocks general-purpose modelling. *arXiv*. 2023
13. Vakhrushev A., Ryzhkov A., Savchenko M. et al. LightAutoML: AutoML solution for a large financial services ecosystem. *arXiv*. 2021
14. Steinegger M., Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol.* 2017;35(11):1026-1028
15. Potter S.C., Luciani A., Eddy S.R. et al. HMMER web server: 2018 update. *Nucleic Acids Res.* 2018;46(W1):W200-W204
16. Mursalim M.K.N., Mengko T.L.E.R., Hertadi R. et al. BiCaps-DBP: Predicting DNA-binding proteins from protein sequences using Bi-LSTM and a 1D-capsule network. *Comput Biol Med.* 2023;163:107241
17. Wei L., Tang J., Zou Q. Local-DPP: An improved DNA-binding protein prediction method by exploring local evolutionary information. *Inf Sci.* 2017;384:135-144

doi 10.18699/bgrs2024-12.3-07

Децентрализованный алгоритм выравнивания нуклеотидных последовательностей

Дедок В.
Институт математики СО РАН, Новосибирск, Россия
dedok@math.nsc.ru

Ключевые слова: сходство строк; выравнивание последовательностей; параллельные вычисления

Мотивация и цель: Среди задач фундаментальной биоинформатики можно выделить задачу выравнивания нуклеотидных последовательностей. Решение этой задачи требуется в филогенетическом анализе, нахождении эволюционно консервативных участков ДНК, но она крайне сложная и входит в класс NP-сложных.

Возможно ли выполнить полный перебор двух последовательностей? Количество способов выравнивания двух последовательностей длины n и m имеет порядок 2^{n+m}, поэтому вычислительно задача полного перебора превращается в решаемую за бесконечное время.

Задача выравнивания последовательностей в биоинформатике является, в каком-то смысле частным случаем задач поиска совпадений в строках для алфавита, состоящего из четырех символов: A, T, G, C. Для решения такой задачи предложено некоторое количество алгоритмов [1, 2]: алгоритм Ратклифа–Мезнера, Смита–Ватермана, Нидлмана–Вунша. Можно выделить группу алгоритмов по выделению длиннейшей общей подпоследовательности LCS, в которых сопоставляются максимально возможное количество элементов заданных строк с сохранением порядка следования.

Очевидным образом напрашивается идея использования суперкомпьютеров. Известны параллельные реализации алгоритмов Нидлмана-Вунш и Matt (multiple alignment with translation and twists, множественное выравнивание с поворотами и сдвигами) – алгоритм, основанный на методе выравненных пар участков структур (AFP) [3–5]. В работе [3] реализация параллельного алгоритма проводилось на суперкомпьютере «Ломоносов», входящем в состав суперкомпьютерного комплекса МГУ имени М.В. Ломоносова.

Использование суперкомпьютеров такого класса – достаточно дорогое решение поэтому авторами была предложена идея использовать децентрализованное решение, основанное на подходе таких проектов, как MilkyWay@Home, Einstein@Home, SETI@home и др. Данный подход имеет название добровольных вычислений – распределенных вычислений с использованием предоставленных добровольно вычислительных ресурсов. Общая схема участия в таком проекте распределенных вычислений состоит в том, что потенциальный участник загружает клиентскую часть программного кода на свой персональный компьютер и запускает ее. Клиентское приложение по сети интернет обращается к серверу, получает вычислительное задание, данные для обработки, выполняет вычисления и их результаты отправляет обратно. При этом вычисления проводятся в фоновом режиме и с минимальным приоритетом, так что работа такого приложения не сказывается на основной работе пользователя. Такой подход был предложен в

середине 1990-х гг. в связи с бурным развитием сети интернет и лавинообразным увеличением количества подключенных к ней компьютеров пользователей.
Недавним примером успехов подобных методов решения тяжелых вычислительных задач является открытие 5000 околосолнечной кометы космической лабораторией SOHO [6] добровольцами гражданской науки Sungrazer [7]. Поэтому авторы пошли по пути разработки алгоритма, позволяющего выполнять биологические вычисления на персональных компьютерах научных волонтеров.
Методы и алгоритмы: В качестве алгоритмов выравнивания последовательностей были выбраны 2 алгоритма: Matt [5] и разработанный авторами итерационный алгоритм последовательного поиска расширяющейся последовательности. Суть предложенного алгоритма можно описать следующей схемой. Пусть у нас имеется две последовательности U и D. В данных последовательностях имеются уже выровненные подпоследовательности U_0 и D_0. Последовательности U и D можно представить в виде $U = U_L U_0 U_R$, $D = D_L D_0 D_R$. На каждом шаге алгоритма будем пытаться расширить имеющиеся последовательности U_0 и D_0 символами из U_L и D_L слева и U_R и D_R справа. В случае, если это возможно, выровненные подпоследовательности будут соответственно расширяться, а расширенные последовательности будут сохраняться в памяти для дальнейшей попытки расширения.
Для распараллеливания вычислений на основе алгоритма Matt используется метод коллективного решения, предложенный в работе [3] и адаптированный для вычислений на отдельных вычислительных узлах.
Мастер-процесс (серверный процесс) на первом этапе алгоритма получает всевозможные пары структур и равномерно распределяет их между всеми доступными вычислительными узлами. Далее происходит выравнивание всеми процессами и сбор данных в мастер-процессе.
Результаты: Результатом данной работы является компьютерная реализация алгоритмов Matt и последовательного поиска расширяющейся последовательности. Клиентская часть обоих алгоритмов реализована в виде хранителя экрана, выполняющегося в периоды неактивности рабочих месте. Серверная часть реализована в виде приложения, работающего в среде ОС Linux. Численные эксперименты показали большую эффективность подхода, основанного на алгоритме Matt в случае, когда одновременно работает достаточно много клиентских приложений и они готовы оперативно принимать вычислительные задания. При этом, если вычислительные узлы с клиентскими приложениями появляются нерегулярно, более эффективным является подход расширяющейся последовательности.
Выводы: Подводя итог исследованию, можно с уверенностью сказать, что подход, основанный на добровольных вычислениях, вполне имеет право на жизнь и позволяет обрабатывать последовательности длиной в десятки тысяч символов на обычных персональных компьютерах научных волонтеров, не требуя при этом мощностей суперкомпьютеров. При этом результирующий алгоритм должен предусматривать этап верификации результатов, полученных на внешних вычислительных узлах, но эта задача на порядки более простая и может быть легко выполнена и не очень производительных ресурсах.

Но при этом стоит отметить и минус такого алгоритма – очень сложно гарантировать время его работы. Но при достаточном количестве пользователей системы этот недостаток может быть практически сведен на нет.
Финансирование: Работа выполнена при финансовой поддержке программы фундаментальных научных исследований СО РАН № I.1.5 (проект FWNF-2022-0009).

Decentralized algorithm for nucleotide sequence alignment

Dedok V.
S.L. Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia
** dedok@math.nsc.ru*

Key words: string similarity; sequence alignment; parallel computing

Motivation and Aim: Among the problems of fundamental bioinformatics, one can highlight the task of aligning nucleotide sequences. The solution to this problem is required in phylogenetic analysis, finding evolutionarily conserved DNA regions, but it is extremely complex and belongs to the class of NP-difficult.
Is it possible to perform an exhaustive search between two sequences? The number of ways to align two sequences of length n and m is of the order of 2^{n+m}, so computationally the exhaustive search problem becomes solvable in infinite time.
The problem of sequence alignment in bioinformatics is, in a sense, a special case of the problem of finding matches in strings for an alphabet consisting of 4 characters A, T, G, C. To solve this problem, a few algorithms have been proposed [1, 2]: Ratcliffe–Mezner, Smith–Waterman, Needleman–Wunsch algorithm. We can distinguish a group of algorithms for identifying the longest common LCS subsequence, which compare the maximum possible number of elements of given strings while maintaining the order.
The obvious idea is to use supercomputers. There are known parallel implementations of the Needleman-Wunsch and Matt algorithms (multiple alignment with translation and twists), an algorithm based on the method of aligned pairs of structure sections (AFP) [3–5]. In work [3], the implementation of the parallel algorithm was carried out on the Lomonosov supercomputer, which is part of the supercomputer complex of Moscow State University named after M.V. Lomonosov.
The use of supercomputers of this class is a rather expensive solution, so the authors proposed the idea of using a decentralized solution based on the approach of projects such as MilkyWay@Home, Einstein@Home, SETI@home, etc. This approach is called voluntary computing - distributed computing using provided volunteer computing resources. The general scheme of participation in such a distributed computing project is that the potential participant downloads the client part of the program code onto his personal computer and runs it. The client application accesses the server over the Internet, receives a computational task, data for processing, performs calculations and sends their results back. In this case, calculations are carried out in the background and with minimal priority, so that the operation of such an application does not affect the user's main work. This approach was proposed in the mid-1990s. due to the rapid development of the Internet and an avalanche-like increase in the number of user computers connected to it.

A recent example of the success of such methods for solving heavy computational problems is the discovery of the 5000th circumsolar comet by the SOHO space laboratory [6] by Sungrazer citizen science volunteers [7]. Therefore, the authors took the path of developing an algorithm that allows performing biological calculations on the personal computers of scientific volunteers.

Methods and Algorithms: Two algorithms were chosen as sequence alignment algorithms: Matt [5] and an iterative algorithm developed by the authors for sequential search for an expanding sequence. The essence of the proposed algorithm can be described by the following diagram. Let us have two sequences U and D. These sequences contain already aligned subsequences U_0 and D_0. Sequences U and D can be represented as $U = U_L U_0 U_R$, $D = D_L D_0 D_R$. At each step of the algorithm, we will try to expand the existing sequences U_0 and D_0 with symbols from U_L and D_L on the left and U_R and D_R on the right. If possible, the aligned subsequences will be expanded accordingly, and the expanded sequences will be stored in memory for further expansion attempts. To parallelize calculations based on the Matt algorithm, the collective solution method proposed in [3] and adapted for calculations on individual computing nodes is used. At the first stage of the algorithm, the master process (server process) receives all possible pairs of structures and distributes them evenly among all available computing nodes. Next, alignment occurs by all processes and data is collected in the master process.

Results: The result of this work is a computer implementation of the Matt algorithm and the sequential search for an expanding sequence. The client part of both algorithms is implemented in the form of a screen saver that runs during periods of inactivity on the desktop. The server part is implemented as an application running in the Linux OS environment.

Numerical experiments have shown the great efficiency of the approach based on the Matt algorithm in the case when quite a lot of client applications are running simultaneously, and they are ready to quickly accept computational tasks. However, if computing nodes with client applications appear irregularly, the expanding sequence approach is more effective.

Conclusion: Summing up the study, we can say with confidence that the approach based on voluntary computing has the right to life and makes it possible to process sequences of tens of thousands of characters in length on ordinary personal computers of scientific volunteers, without requiring the power of supercomputers. In this case, the resulting algorithm must include the stage of verifying the results obtained on external computing nodes, but this task is orders of magnitude simpler and can be easily completed even with not very productive resources. But it is worth noting the disadvantage of such an algorithm – it is very difficult to guarantee its operating time. But with a sufficient number of clients, this disadvantage can be practically eliminated.

Funding: The study is supported by the SB RAS program of fundamental scientific research No. I.1.5 (project FWNF-2022-0009).

Список литературы/References

1. Durbin R., Eddy S., Krogh A., Mitchison G. Biological sequence analysis: probabilistic models of proteins and nucleic acidsю Cambridge University Press, 1998
2. Знаменский С.В. Модель и алгоритм выравнивания последовательностей. *Программные системы: теория и приложения*. 2015;1(24):189-197
   [Znamenskij S.V. A model and algorithm for sequence alignment. *Program Systems: Theory and Applications*. 2015;1(24):189-197 (in Russian)]

3. Воеводин В.В., Шегай М.В., Попова Н.Н., Суплатов Д.А., Швядас В.К. Использование параллельных вычислений для эффективного построения множественных выравниваний структур белков. В: Параллельные вычислительные технологии (ПаВТ'2016). Архангельск: Издательский центр ЮУрГУ, 2016;81-92
[Voevodin V.V., Shegay M.V., Popova N.N., Suplatov D.A., Svedas V.K. Use of parallel computing in effective protein multiple structure alignment. In: Parallel computational technologies (PCT'2016). Arkhangelsk, 2016;81-92 (in Russian)]
4. Тетуев Р.К., Пятков М.И., Панкратов А.Н. Параллельный алгоритм глобального выравнивания протяжённых аминокислотных и нуклеотидных последовательностей. *Математическая биология и биоинформатика*. 2017;12(1):137-150. doi 10.17537/2017.12.137
[Tetuev R.K., Pyatkov M.I., Pankratov A.N. Parallel algorithm for global alignment of long aminoacid and nucleotide sequences. *Math Biol Bioinform*. 2017;12(1):137-150. doi 10.17537/2017.12.137 (in Russian)]
5. Menke M., Berger B., Cowen L. (2008) Matt: Local Flexibility Aids Protein Multiple Structure Alignment. *PLoS Comput Biol*. 2008;4(1):e10. doi 10.1371/journal.pcbi.0040010
6. ESA, NASA Solar Observatory Discovers Its 5,000th Comet. https://science.nasa.gov/science-research/heliophysics/esa-nasa-solar-observatory-discovers-its-5000th-comet/
7. The Sungrazer Project. https://science.nasa.gov/citizen-science/the-sungrazer-project/

doi 10.18699/bgrs2024-12.3-08

TCRscape – инструмент для мультимодального кластерирования единичных Т-клеток на основе последовательностей их Т-клеточных рецепторов и экспрессии ключевых генов

Перик-Заводский Р., Перик-Заводская О., Волынец М., Алрхмун С., Сенников С.*

Научно-исследовательский институт фундаментальной и клинической иммунологии (НИИФКИ) Министерства науки и высшего образования Российской Федерации, Новосибирск, Россия
** sennikov@niikim.ru*

Ключевые слова: Т-клетки; секвенирование Т-клеточного рецептора единичных клеток; секвенирование РНК единичных клеток; клонотипирование Т-клеток; транскриптомика; BD Rhapsody

Мотивация и цель: Целью работы является создание инструмента на основе метода кластерирования единичных клеток, способного одновременно определять доминантные клонотипы и оценивать транскриптом антиген-специфических TCR Т-клеток, полученных при помощи инструмента BD Rhapsody.
Методы и алгоритмы: TCRscape использует матрицы экспрессии генов единичных клеток и матрицу репертуара рецепторов адаптивного иммунитета (AIRR), сгенерированные инструментом для обработки сырых данных секвенирования РНК единичных клеток BD Rhapsody. Для исследуемой матрицы экспрессии генов выполняется log2(Counts Per Million) нормализация, а с использованием матрицы AIRR обнаруживаются и подсчитываются клонотипы Т-клеточных рецепторов (TCR). Избранные гены нормализованной матрицы экспрессии генов, связанные со специфичностью и аффинностью TCR, объединяются с матрицей клонотипов, далее для объединенной матрицы выполняется анализ главных компонентов (PCA) с целью оценки размерности данных, выполняется уменьшение размерности Uniform Manifold Approximation and Projection (UMAP) [1] с учетом размерности данных, а затем ищутся кластеры с использованием метода HDBSCAN [2] (рис. 1).
Результаты: Нами был создан инструмент, способный в течение нескольких минут обнаружить доминантные клонотипы и оценить фенотип антиген-специфических TCR Т-клеток, полученных при помощи инструмента BD Rhapsody.
Выводы: TCRscape является удобным и быстрым инструментом для поиска клонотипов Т-клеточных рецепторов для выполнения обширного списка последующих задач. Выбранные клонотипы Т-клеточных рецепторов будут являться кандидатами для создания терапевтических конструкций, которые могут быть использованы во многих областях, включая лечение неопластических заболеваний и модуляцию иммунных реакций с помощью Treg-клеток.
Финансирование: Работа выполнена при поддержке Российского научного фонда № 21-65-00004 (https://rscf.ru/project/21-65-00004).

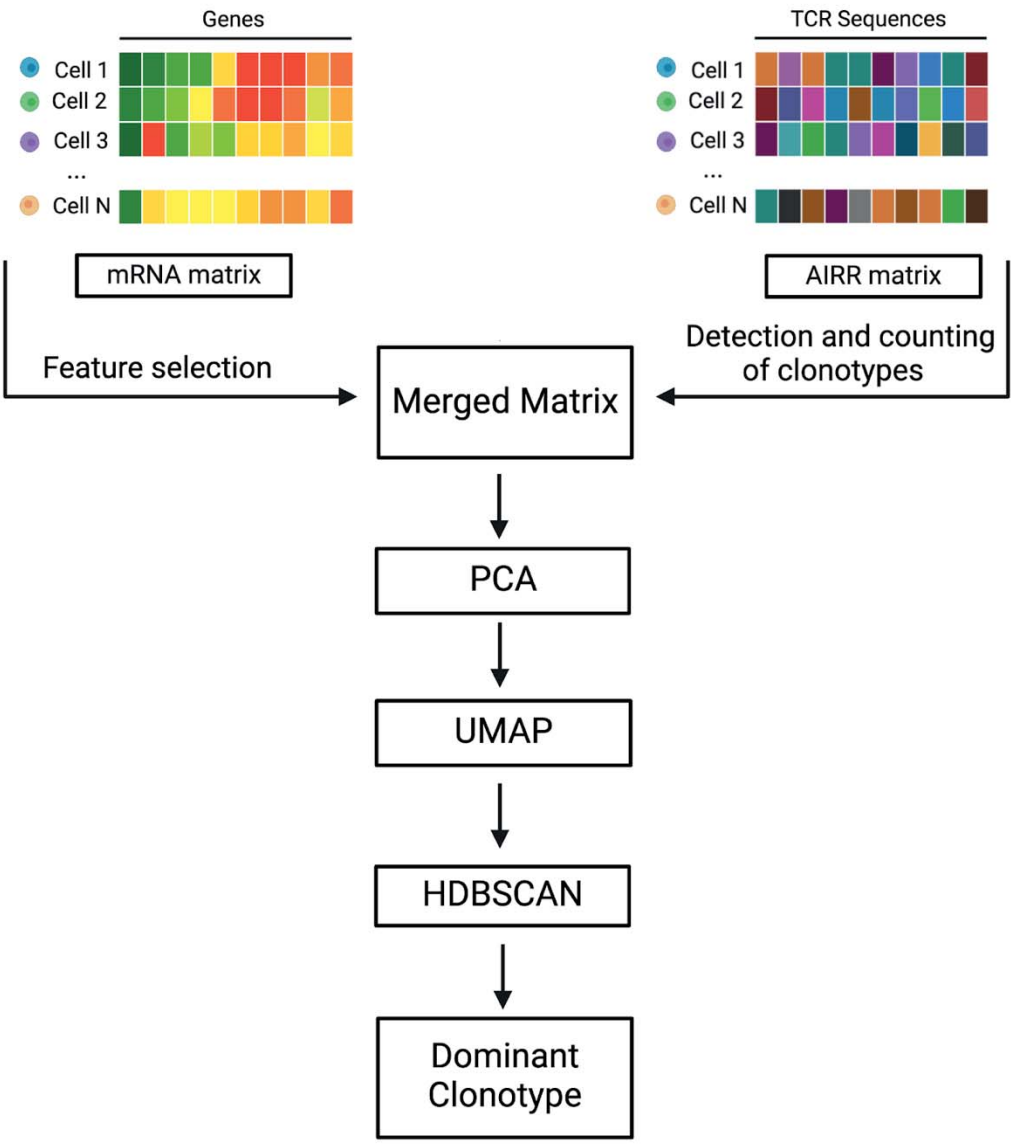


Рис. 1. Схема работы инструмента TCRscape

TCRscape – a tool for simultaneous multimodal gene expression analysis of single T-cells and their clonotypes

Perik-Zavodskii R., Perik-Zavodskaia O., Volynets M., Alrhmoun S., Sennikov S.*
Laboratory of molecular immunology, Research Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia
** sennikov@niikim.ru*

Key words: T-cells; scTCR-seq; scRNA-seq; T-cell clonotyping; transcriptomics; BD Rhapsody

Motivation and Aim: The aim of this work is to create a tool capable with the help of single cell clustering methods to identify dominant clonotypes and evaluate the phenotype of antigen-specific TCR T cells obtained using the BD Rhapsody single-cell analysis system.

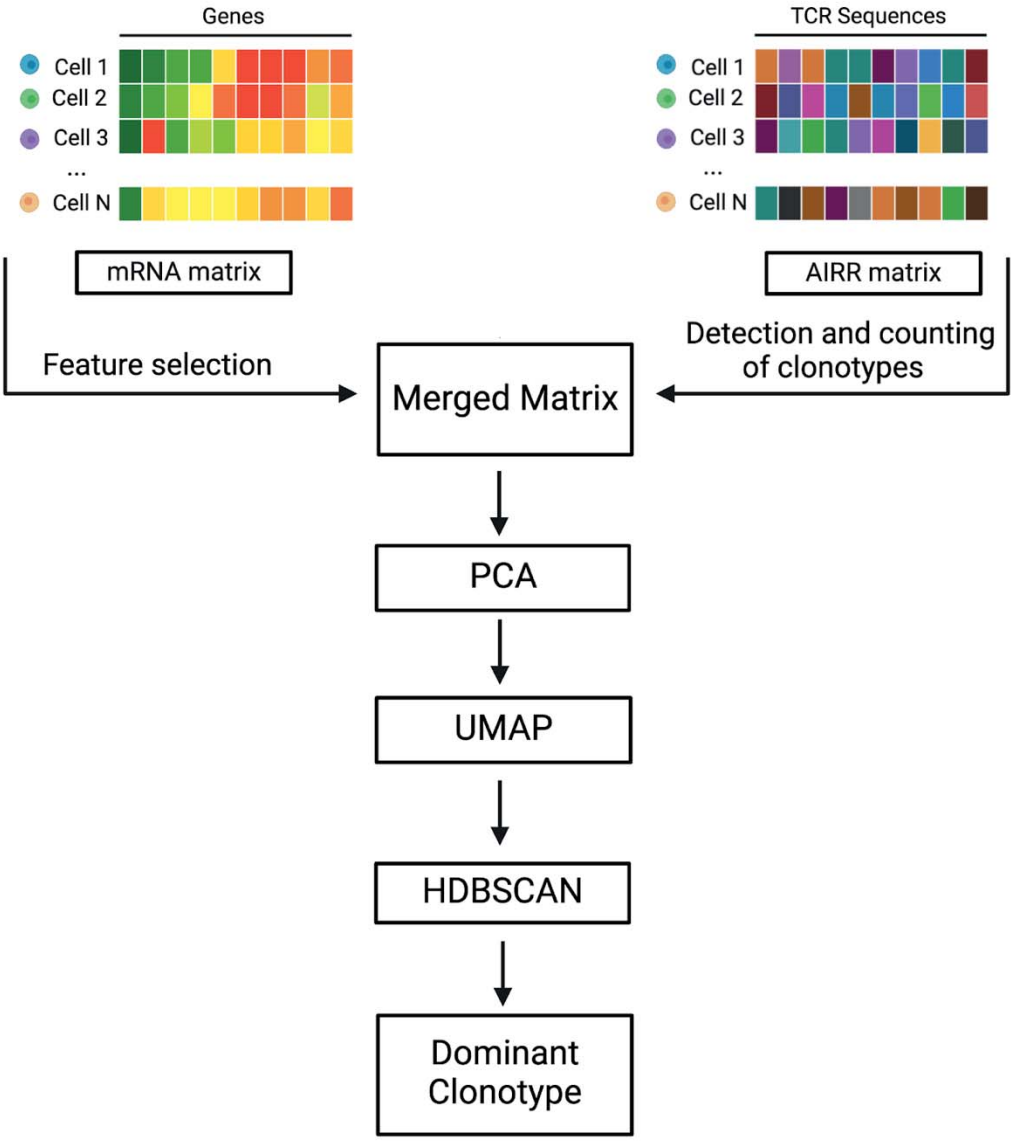


Fig. 1. TCRscape tool

Methods and Algorithms: TCRscape uses the single cell gene expression matrix and the Adaptive Immune Receptor Repertoire (AIRR) matrix generated by the BD Rhapsody pipeline for the processing of raw RNA sequencing data. Log2(Counts Per Million) normalization is then performed on the studied gene expression matrices, and using the AIRR matrix, clonotypes of T cell receptors (TCR) are detected and counted; after that, genes associated with TCR specificity and affinity are picked from the normalized gene expression matrix, which are then combined with the clonotypes matrix, subsequently, a principal component analysis (PCA) is performed on the combined matrix in order to

assess the dimensionality of the data, followed by performing Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction [1] taking into account the dimensionality of the data from the PCA analysis, finally, clusters are identified using the HDBSCAN method [2] (Fig. 1).

Results: We have created a tool that can simultaneously identify dominant clonotypes and assess the phenotype of antigen-specific TCR T-cells obtained using the BD Rhapsody instrument within minutes.

Conclusion: TCRscape is a convenient and fast tool for screening T-cell receptor clonotypes for a comprehensive list of downstream tasks. The selected T-cell receptor clonotypes can serve as candidates for the development of therapeutic constructs that have the potential to be used in many areas, including the treatment of neoplastic diseases and the modulation of immune responses by Treg cells.

Funding: This work was carried out with the support of the Russian Science Foundation, project No. 21-65-00004 (https://rscf.ru/project/21-65-00004).

Список литературы/References

1. McInnes L., Healy J., Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv*. 2018. doi 10.21105/joss.00861
2. McInnes L., Healy J., Astels S. hdbscan: Hierarchical density based clustering. *J Open Source Softw*. 2017;2(11):205. doi 10.21105/joss.00205

doi 10.18699/bgrs2024-12.3-09

Бинарно-геномные числа и симметрии универсальных правил статистической организации геномных ДНК

Петухов С.В.
Институт машиноведения им. А.А. Благонравова РАН, Москва, Россия
spetoukhov@gmail.com

Ключевые слова: геномные ДНК; биоинформатика; правила вероятностей; многоуровневые тексты; дихотомии; симметрии; фрактальные сети

Мотивация и цель: Создатели квантовой механики П. Йордан и Э. Шредингер указывали на ключевое отличие живых тел от неодушевленных: неодушевленные объекты управляются средним случайным движением их миллионов частиц и движение отдельных частиц не существенно для целого; напротив, в живом организме избранные – генетические – молекулы обладают диктаторским влиянием на весь организм за счет квантового усиления (см. историю «квантовой биологии» [1]). Кроме того, Йордан утверждал, что упущенные наукой законы жизни являются законами вероятностей квантового мира. Целью представляемого авторского исследования является поиск этих упущенных законов жизни путем анализа статистических особенностей информационных последовательностей нуклеотидов в однонитевых ДНК геномов высших и низших организмов из общедоступного банка генетических данных GenBank.
Методы и алгоритмы: Авторский метод анализа статистических закономерностей последовательности нуклеотидов любой из однонитевых геномных ДНК заключается в ее бинарном представлении как последовательности нуклеотидов двух видов: пуринов (A и G), обозначаемых символом 0, и пиримидинов (C и T), обозначаемых символом 1. В докладе этот метод поясняется примером информационной последовательности ДНК хромосомы человека № 1, содержащей около 250 миллионов нуклеотидов. Ее бинарное (пурин-пиримидиновое) представление имеет форму сверхогромного бинарного числа с около 250 миллионами битов (десятичный аналог такого числа доходит до 2^250000000). Эти числа названы докладчиком «бинарно-геномными» (БГ-числами). В них предположительно отражены законы квантовой биофизики и квантовой информатики, генетическая память и пр. Математическое естествознание прежде не работало с системами таких огромных чисел (насколько докладчику известно). Для обнаружения скрытых в БГ-числах статистических закономерностей они анализировались авторским алгоритмом «статистических матрешек» (алгоритм иерархии бинарных статистик), включающим следующие стадии анализа процентного состава БГ-числа: сначала оно представляется как последовательность одиночных символов 0-1-1-1-0-0-… и в ней подсчитываются %0 и %1; затем оно представляется как последовательность дуплетов 01-11-00-10-… и подсчитываются проценты каждого вида бинарных дуплетов %00, %01, %10, %11; затем аналогично та же бинарная последовательность представляется как последовательность триплетов, тетраплетов, пентаплетов, …, с подсчетом каждый раз процентов соответствующих видов n-плетов. Тем самым каждое БГ-число представляется как многослойное, каждый слой которого написан алфавитом

соответствующей диадической группы n-битных чисел. В итоге для БГ-числа получаем таблицу вероятностей каждого вида n-плетов в соответствующем n-плетном слое.

Результаты: Автором выявлено, что в таких таблицах вероятностей для БГ-чисел между процентами n-плетов, а также между их семействами при разных n имеются нетривиальные связи. Рисунок 1 показывает пример зеркальной симметрии распределения вероятностей в исследованных семействах бинарных дуплетов, триплетов, тетраплетов и пентаплетов для случая хромосомы человека № 1. Красным обозначены проценты n-плетов, начинающихся с 0, а синим – с 1. Взаимозамена 0↔1 дает пары комплементарных n-плетов равной вероятности.

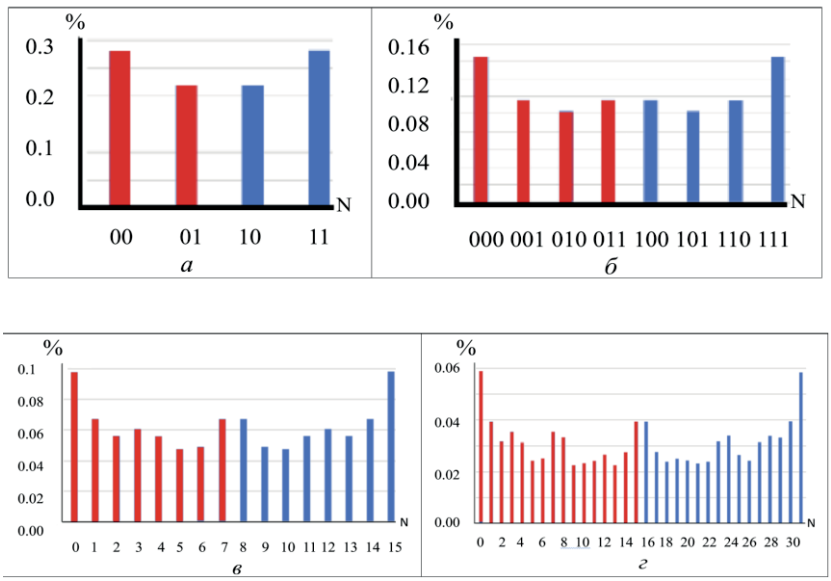


Рис. 1. Распределение вероятностей в бинарных семействах 4 дуплетов, 8 триплетов, 16 тетраплетов и 32 пентаплетов для БГ-числа пурин-пиримидиновой последовательности однонитевой ДНК хромосомы человека № 1

Наряду с этим в многослойной статистике данного бинарно-геномного числа между процентами n-плетов пуринов и пиримидинов из семейств с разными значениями n имеется высокоточная дихотомическая связь (при изученных значениях n = 1, 2, 3, 4, 5): процент любого n-плета практически равен сумме процентов таких двух (n+1)-плетов, которые отличаются от него наличием суффиксов 0 и 1. Рисунок 2 показывает – для БГ-числа ДНК хромосомы человека № 1 – пример этой связи в форме диаграммы дихотомического характера, а также соответствующего фрактального дихотомического дерева процентов n-плетов, в котором каждый узел является началом своего собственного фрактального дихотомического дерева вероятностей.

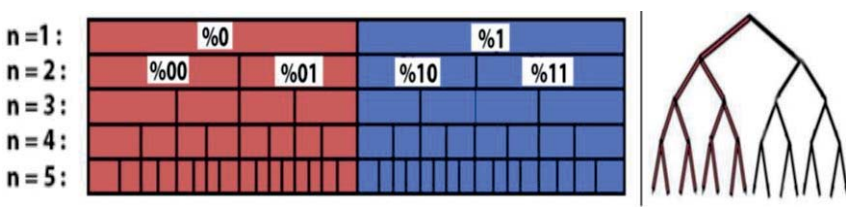


Рис. 2. Диаграмма дихотомических связей описываемых вероятностей

Это правило дихотомий в статистической организации БГ-чисел справедливо для всех уже исследованных нами геномных ДНК: всех 24 хромосом человека; всех хромосом дрозофилы, мыши, червя, многих растений; 19 геномов бактерий и архей; многих экстремофилов, живущих в экстремальных условиях, например, радиации с уровнем, в 1000 раз превышающим смертельный для человека [2–4].
Тем самым подтверждается предвидение Йордана о существовании упущенных наукой законов жизни, являющихся законами вероятностей квантового мира. Аналогично Шредингер в книге [5] утверждал: «*Для организма следует ожидать новых законов. ... все известное нам о структуре живой материи заставляет ожидать, что деятельность живого организма нельзя свести к проявлению обычных законов физики ... Мы вправе предполагать, что живая материя подчиняется новому типу физического закона*». Докладываемые общие геномные правила – кандидаты на роль таких законов. Эти универсальные правила геномных ДНК говорят о существовании нетривиальных алгебраических инвариантов глобального геномного характера, которые остаются неизменными на протяжении миллиардов лет эволюции, в ходе которой отмирают миллионы видов организмов и возникают новые (хотя локально геномные последовательности изменяются мутациями, прессом естественного отбора и пр.).
Дихотомии в генетически наследуемых телах хорошо известны: два полушария мозга; дихотомические ветвления у растений, сосудов и нейронов; бронхиальное дерево легких человека и пр. Но в геномных ДНК, в отличие от телесных конфигураций, мы сталкиваемся с принципиально иным типом дихотомии: дихотомиями вероятностных характеристик в информационных последовательностях ДНК. Обширные дихотомические фрактальные сети вероятностей геномных ДНК – это та почва, из которой растут живые тела и генетический интеллект. Материальные структуры живых тел не возникают из ниоткуда, а имеют структурные прототипы в регулярной системе геномных вероятностей. Автор полагает, что существует скрытый от непосредственного восприятия мир семейств вероятностей, структурированных на основе бинарных оппозиций. Именно по образу и подобию бинарно организованных семейств вероятностей этого многомерного мира строятся генетически наследуемые биологические тела. Наши тела являются как бы одеждой, надетой на эти бинарно структурированные семейства вероятностей, выступающие прототипами биологических тел. Напомним, что генетика как наука началась с открытия Менделем статистических правил наследования признаков. Многие процессы в живых организмах стохастические, поскольку отдельные молекулы взаимодействуют в клетках стохастическим образом.
Выводы: Существуют универсальные правила статистической организации информационных последовательностей геномных ДНК высших и низших организмов. Они сопряжены со статистическими закономерностями соответствующих сверхогромных бинарно-геномных чисел, представляющих собой новый класс бинарных чисел математического естествознания. Полученные результаты затрагивают многие вопросы биоинформатики и создания искусственного интеллекта геноморфного типа [2–4].

Binary-genomic numbers and symmetries of the universal rules of statistical organization of genomic DNAs

Petoukhov S.V.
Mechanical Engineering Research Institute named after A.A. Blagonravov, RAS, Moscow, Russia
spetoukhov@gmail.com

Key words: genomic DNA; bioinformatics; probability rules; multi-level texts; dichotomies; symmetries; fractal networks

Motivation and Aim: The creators of quantum mechanics, P. Jordan and E. Schrödinger, pointed out the key difference between living bodies and inanimate ones: inanimate objects are controlled by the average random movement of their millions of particles and the movement of individual particles is not significant for the whole; on the contrary, in a living organism, selected – genetic – molecules have a dictatorial influence on the entire organism due to quantum enhancement (see the history of "quantum biology" [1]). In addition, Jordan argued that the laws of life missed by science are the laws of probability of the quantum world. The purpose of the present author's research is to search for these missed laws of life by analyzing the statistical features of information sequences of nucleotides in single-stranded DNA genomes of higher and lower organisms from the publicly available genetic data bank GenBank.
Methods and Algorithms: The author's method for analyzing statistical patterns in the nucleotide sequence of any single-stranded genomic DNA consists in its binary representation as a sequence of two types of nucleotides: purines (A and G), designated by the symbol 0, and pyrimidines (C and T), designated by the symbol 1. In the report, this The method is illustrated by the example of the DNA information sequence of human chromosome No. 1, which contains about 250 million nucleotides. Its binary (purine-pyrimidine) representation has the form of a super-huge binary number with about 250 million bits (the decimal analogue of such a number reaches 2^250,000,000). The speaker called these numbers "binary-genomic" (BG-numbers). They presumably reflect the laws of quantum biophysics and quantum information science, genetic memory, etc. Mathematical natural science has not previously worked with systems of such huge numbers (as far as the speaker knows). To detect statistical patterns hidden in BG-numbers, they were analyzed by the author's "statistical nesting dolls" algorithm (or an algorithm for the hierarchy of binary statistics), which included the following stages of analysis of the percentage composition of a BG-number: first, it is represented as a sequence of single characters 0-1-1-1-0-0-... and %0 and %1 are calculated in it; then it is represented as a sequence of duplets 01-11-00-10-... and the percentages of each type of binary duplets %00, %01, %10, %11 are calculated; then, similarly, the same binary sequence is represented as a sequence of triplets, tetraplets, pentaplets, ..., each time counting the percentage of the corresponding types of n-plets. Thus, each BG-number is represented as a multilayer number, each layer of which is written in the alphabet of the corresponding dyadic group of n-bit numbers. As a result, for the BG-number we obtain a table of probabilities for each type of n-plets in the corresponding n-plet layer.
Results: The author has discovered that in such probability tables for BG-numbers there are non-trivial connections between the percentages of n-plets, as well as between their families for different n. Figure 1 shows an example of mirror symmetry of the probability

distribution in the studied families of binary duplets, triplets, tetraplets and pentaplets for the case of human chromosome No. 1. Red indicates the percentage of n-plets starting with 0, and blue – with 1. The 0↔1 interchange reveals pairs of complementary n-plets of equal probability.

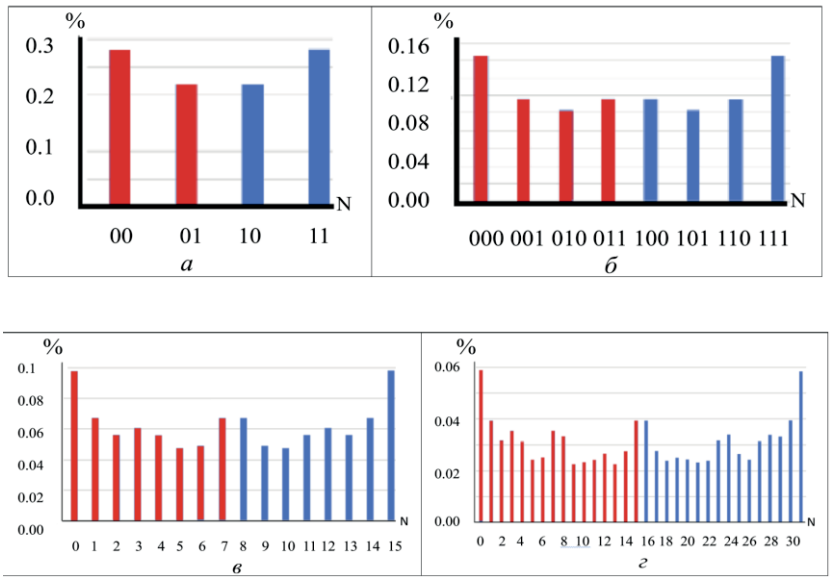


Fig. 1. Probability distribution in binary families of 4 duplets, 8 triplets, 16 tetraplets and 32 pentaplets for the BG-number of the purine-pyrimidine sequence of single-stranded DNA of human chromosome No. 1

Along with this, in the multilayer statistics of a given binary-genomic number, there is a highly accurate dichotomous relationship between the percentages of n-plets of purines and pyrimidines from families with different values of n (for the studied values of n = 1, 2, 3, 4, 5): the percentage of any n- plet is practically equal to the sum of the percentages of two (n+1)-plets that differ from it by the presence of suffixes 0 and 1. Fig. 2 shows – for the BG-number of DNA of human chromosome No. 1 – an example of this connection in the form of a diagram of a dichotomous nature, as well as the corresponding fractal dichotomous tree of n-plet percentages, in which each node is the beginning of its own fractal dichotomous tree of probabilities.

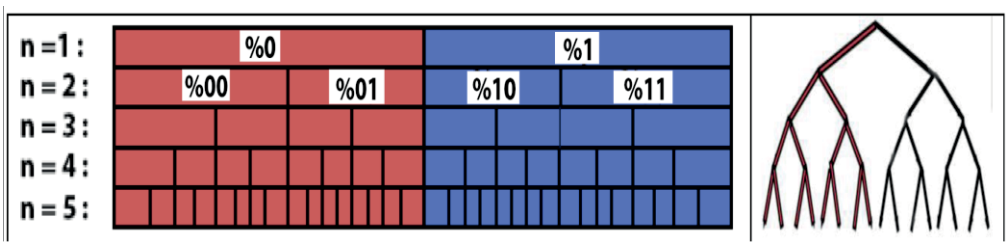


Fig. 2. Diagram of dichotomous connections of the described probabilities

This rule of dichotomies in the statistical organization of BG-numbers is true for all genomic DNA that we have already studied: all 24 human chromosomes; all chromosomes of Drosophila, mouse, worm, many plants; 19 genomes of bacteria and

archaea; many extremophiles living in extreme conditions, such as radiation levels 1,000 times more than lethal to humans [2–4].

Thus, Jordan's prediction about the existence of laws of life missed by science, which are the laws of probabilities of the quantum world, is confirmed.

Similarly, Schrödinger stated: "*New laws to be expected in the organism. What I wish to make clear in this last chapter is, in short, that from all we have learned about the structure of living matter, we must be prepared to find it working in a manner that cannot be reduced to the ordinary laws of physics*" [5]. The reported general genomic rules of probabilities are candidates for the role of such laws.

These universal rules of genomic DNAs indicate the existence of non-trivial algebraic invariants of a global genomic nature, which remain unchanged over billions of years of evolution, during which millions of species of organisms die off and new ones arise (although locally genomic sequences change due to mutations, the pressure of natural selection, etc.).

The dichotomies in genetically inherited bodies are well known: two hemispheres of the brain; dichotomous branching in plants, vessels and neurons; bronchial tree of human lungs; etc. But in genomic DNAs, in contrast to bodily configurations, we are faced with a fundamentally different type of dichotomy: dichotomies of probabilistic characteristics in DNAs information sequences. Vast dichotomous fractal networks of genomic DNAs probabilities are the soil from which living bodies and genetic intelligence grow. The material structures of living bodies do not appear out of nowhere, but have structural prototypes in a regular system of genomic probabilities. The author believes that there is a world of families of probabilities, hidden from direct perception, structured on the basis of binary oppositions. It is in the image and likeness of binary organized families of probabilities of this multidimensional world that genetically inherited biological bodies are built. Our bodies are like clothes put on these binary structured families of probabilities, which act as prototypes of biological bodies. Let us recall that genetics as a science began with Mendel's discovery of the statistical rules of inheritance of traits. Many processes in living organisms are stochastic because individual molecules interact in cells in a stochastic manner.

Conclusion: There are universal rules for the statistical organization of information sequences of genomic DNAs of higher and lower organisms. They are associated with statistical patterns of the corresponding super-huge binary-genomic numbers, which represent a new class of binary numbers in mathematical natural science. The results obtained touch on many issues of bioinformatics and the creation of artificial intelligence of the genomorphic type.

Список литературы/References

1. McFadden J., Al-Khalili J. The origins of quantum biology. *Proc Math Phys Eng Sci.* 2018;474(2220):20180674. doi 10.1098/rspa.2018.0674
2. Petoukhov S.V. Cyclic Gray Codes in Modeling Inherited Cyclic Biostructures and Analysis of Statistical Rules of Genomic DNAs. *Preprints.* 2024. doi 10.20944/preprints202402.0713.v1
3. Petoukhov S.V. The Principle "Like Begets Like" in Molecular and Algebraic-Matrix Genetics. *Preprints.* 2022. doi 10.20944/preprints202211.0528.v3
4. Petoukhov S.V., He M. Algebraic Biology, Matrix Genetics, and Genetic Intelligence. Singapore; World Scientific, 2023. doi 10.1142/13468
5. Schrodinger E. What is Life? The Physical Aspect of the Living Cell. Cambridge University Press, 1955

doi 10.18699/bgrs2024-12.3-10

Машинное обучение в прогнозировании эпизодов застывания на ультраресурсоограниченных вычислительных устройствах

Сейкин А.А.[1]*, Кириллов Б.А.[2, 3]

[1] *Московский физико-технологический институт, Долгопрудный, Россия*

[2] *Центр технологий материалов, Сколковский институт науки и технологий, Москва, Россия*

[3] *Центр высокоточного редактирования и генетических технологий для биомедицины, Институт биологии гена РАН, Москва, Россия*

* *seikin.aa@phystech.edu*

Ключевые слова: tinyML; нейронный сети; двигательные нарушения; болезнь Паркинсона

Мотивация и цель: Застывание, или «замерзание» походки (freezing of gait, FOG) – серьезное осложнение болезни Паркинсона, заключающееся во временной неспособности инициировать процесс ходьбы, значительно влияющее на качество жизни пациентов. Наша исследовательская работа направлена на создание системы, способной предсказывать FOG на основе данных с акселерометров, используя энергоэффективные микроконтроллеры с ограниченными ресурсами. Уникальность нашей системы заключается в ее способности работать в условиях строгих ограничений по питанию и вычислительной мощности, характерных для носимых устройств [1]. Мы исследуем возможность использования различных микроконтроллеров, таких как Cortex-M, ATmega328, ESP8266. Отдельной задачей проекта является разработка в области развертывания компактифицированной модели на микроконтроллерах российского производства «АМУР» на базе архитектуры RISC-V, предлагающих оптимальный баланс между энергопотреблением и вычислительной способностью.

Методы и алгоритмы: Для решения задачи предсказания FOG мы применяем методы генерации сжатых весов обученной модели машинного обучения путем применения ряда фреймворков, таких как Tensorflow Lite и аналоги [2]. Сначала на основе датасета по двигательным нарушениям при болезни Паркинсона [3] мы создаем традиционную модель машинного обучения, затем сериализуем ее, структурно и логически модифицируем для минимизации размера весов, необходимых для предсказания замерзания походки при развертывании модели на микроконтроллерах.

Результаты: В предлагаемой работе мы исследуем возможность прогноза двигательных нарушений на носимых устройствах с применением вычислительных ресурсов микроконтроллера, в условиях экстремального недостатка мощности, и проводим анализ набора микроконтроллеров, доступных на российском рынке с точки зрения точности предсказания и скорости вычислений.

Выводы: С помощью алгоритмов машинного обучения, адаптированных для работ ы в экстремальных ограничениях на вычислительную мощность, наша система анализирует паттерны движения и определяет вероятность возникновения FOG [4]. Предсказание FOG, возможное даже на устройствах с ограниченными ресурсами, открывает путь к разработке доступных, надежных и

энергоэффективных носимых систем для улучшения жизни людей, страдающих болезнью Паркинсона [5].

Machine learning in predicting freezing of gait events using ultra-resource-constrained computational devices

Seikin A.[1]*, Kirillov B.[2, 3]

[1] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

[2] *Center of Material Technologies, Skolkovo Institute of Science and Technology, Moscow, Russia*

[3] *Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, RAS, Moscow, Russia*

** seikin.aa@phystech.edu*

Key words: tinyML; neural networks; movement disorders; Parkinson's disease

Motivation and Aim: Freezing of Gait (FOG), a sudden temporary inability to initiate walking, is a severe complication of Parkinson's disease that significantly impacts patients' quality of life. Our research aims to develop a system that can predict FOG based on accelerometer data, utilizing energy-efficient microcontrollers with limited resources. The uniqueness of our system lies in its ability to operate within strict power and computational constraints, characteristic of wearable devices [1]. We explore the use of various microcontrollers, such as Cortex-M, ATmega328, ESP8266. A separate goal of our project is to develop and deploy compact models on Russian-made “АМУР” microcontrollers that offer an optimal balance between energy consumption and computational power.

Methods and Algorithms: Our primary method involves generating compact representations of trained machine learning models by applying a range of frameworks [2]. Specifically, we create a traditional machine learning model based on a Parkinson's disease motor disorder dataset, then serialize and structurally and logically modify it to minimize the information volumes required for prediction, making it deployable on microcontrollers.

To solve Freezing of Gait prediction, we apply methods to generate compressed weights for an already trained machine learning model using a set of frameworks, such as Tensorflow Lite and analogs. First, based on a Parkinson-related movement disorder dataset [3], we train a traditional deep learning model, then we serialize it, modify its structure and logic in order to minimize the size of its weights. This is needed to predict FOG on a microcontroller.

Results: In this research, we study the possibility of predicting motor disorders on wearable devices using microcontroller computational resources in an extremely compute-starved environment and analyze a set of microcontrollers available on Russian market according to the prediction quality of microcontroller-adapted models and the speed of their inference.

Conclusions: By employing machine learning algorithms adapted for operation within extreme constraints on computational power, our system analyzes movement patterns and determines the likelihood of FOG occurrence [4]. FOG prediction, enabled on resource-constrained devices, paves the way for the development of affordable, reliable, and energy-efficient wearable systems to improve the lives of people with Parkinson's disease [5].

Список литературы/References

1. Gupta A.K., Nandyala S.P. Deep Learning on Microcontrollers: Learn How to Develop Embedded AI Applications Using TinyML. BPB Publications, 2023
2. Warden P., Situnayake D. TinyML Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers. O'Reilly Media, 2019
3. Zach H. et al. Identifying freezing of gait in Parkinson's disease during freezing provoking tasks using waist-mounted accelerometry. *Parkinsonism Relat Disord*. 2015;21(11):1362-1366
4. Hou Y. et al. Multi-Modal Wireless Flexible Gel-Free Sensors with Edge Deep Learning for Detecting and Alerting Freezing of Gait in Parkinson's Patients. *arXiv*. 2023
5. Prakash S. et al. Is TinyML Sustainable? *Commun ACM*. 2023;66(11):68-77

doi 10.18699/bgrs2024-12.3-11

Интеллектуальная эпидемия в науке о миРНК

Фирсов А.[1]*, Титов И.[2, 3]

[1] *Институт систем информатики им. А.П. Ершова СО РАН, Новосибирск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

[3] *Новосибирский государственный университет, Новосибирск, Россия*

* *artemijfirsov@mail.ru*

Key words: *k*-мер; *n*-gram; dbscan; идентификация; микроРНК; timsort; kofer; электронная библиотека; соавторство организаций; малый мир; степенной рост

Мотивация и цель: Цифровая форма научных статей позволяет автоматически запрашивать их метаданные, которые могут быть использованы для оценки статистики научной области. В нашей работе мы сосредоточились на области науки о миРНК, для которой мы построили сеть соавторства организаций, чтобы обеспечить анализ на более высоком уровне [1] эволюции научного сообщества по сравнению с характеристиками авторов. В данной работе мы показываем, что развитие области науки о миРНК описывается эпидемиологической моделью [2], которая расширяет распространение модели SIR на трансмиссионный график. Ранее мы показали, что рост области миРНК подчиняется архитектуре малого мира [3]. В этой работе мы предоставляем анализ степенного роста публикаций в этой области, а также показываем сравнение на уровне стран, выделяя США и Китай как лидеров.

Распространение знания как эпидемия: Болезни распространяются, когда инфекционный материал встречает восприимчивое население, состоящее из Susceptible, Infectious, or Recovered. Воздействие может привести к иммунитету или инфицированию с латентным периодом до появления симптомов. Эпидемии возникают, когда наблюдается всплеск заболеваний. Гофман [4] расширил это, обсуждая эпидемии идей, такие как религиозные или научные концепции. В отличие от желательных эпидемий идей, вспышки заболеваний нежелательны. В научных или лаконичных населениях уязвимые члены могут заразиться концепцией, в то время как иммунные индивиды могут отвергнуть ее или распространять ее ненамеренно. Когда значительная часть населения принимает концепцию, происходит эпидемия знаний.

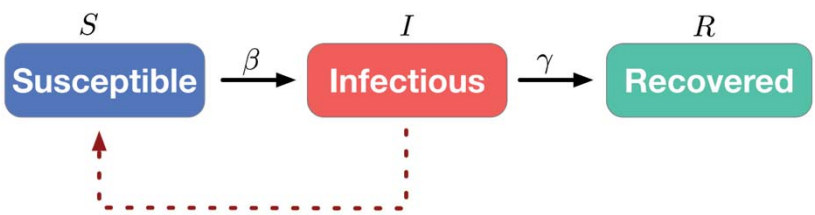


Рис. 1. Переход от восприимчивого к инфицированному и к выздоровевшему состоянию в модели SIR

Такая концепция была применена к популяции с известным трансмиссионным графом в [2], что обсуждается в следующем разделе.

Динамика сетей малого мира: Распространение знаний можно сравнить с эпидемией, где организации уязвимы для «заражения» определенными научными областями и впоследствии публикуются в них. Эти динамики часто представляются с использованием моделей Susceptible, Infectious, or Recovered (SIR) [6], которые демонстрируют экспоненциальный рост инфекций на начальных стадиях. Тем не менее эти модели не учитывают основной граф передачи заболевания. В [2] автор моделирует рост заболеваемости для проблем, где известны графы передачи и имеется свойство малого мира [7], более подробное объяснение было дано в [3], и полученная модель для «зараженных» организаций:

$$\rho(t) = \lambda \frac{(\lambda t)^{D-1}}{(D-1)!} e^{-(\lambda-\mu)t} \left[1 + O\left(\frac{t_0}{t}\right)\right],$$

где $\lambda/\ \mu$ – параметры экспоненциации заражения/выздоровления из модели SIR, D – средний кратчайший путь графа, t_0 время переключения между степенным ростом и замедлением экспоненты.

Методы и алгоритмы: Наш подход включал использование цифровой библиотеки PubMed для компиляции набора данных по исследованиям миРНК. Аффиляции, упомянутые в этих статьях, были сгруппированы с помощью лексикографической сортировки предварительно нормализованных и разделенных векторов признаков K-mer булева типа. Эти сгруппированные аффиляции содержат ссылки на уникальные организации, что облегчает построение графа соавторства для выявления организаций, сотрудничавших при публикациях.

Результаты: Как было показано ранее [3], граф миРНК обладает свойством малого мира, и с $D = 3.46$, и аппроксимированным параметром экспоненциации $D - 1 \approx 2.6$, мы видим отклонение на 5.5 % от прогноза модели.

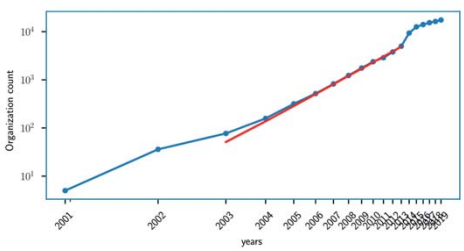


Рис. 2. Ежегодно учитываются организации, опубликовавшие статью в течение определенного года

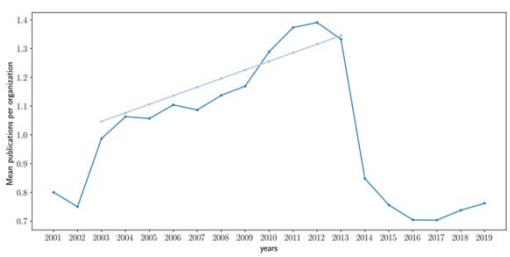


Рис. 3. Среднее количество публикаций на организацию в год

В данной работе мы показываем, что среднее количество публикаций на организацию следует линейному росту, который может быть аппроксимирован и использован для сравнения точной динамики публикаций с оцененной с использованием модели степенного закона, а также динамики среднего количества публикаций на организацию:

$$a_{org} * x_{org}^{b_{org}} * (mx + l) = a * t^b,$$

где $mx + l$ – среднее количество публикаций на организацию в год, b_{org} – параметр экспоненциации степенного роста организаций, b – параметр экспоненциации степенного роста публикаций, a_{org}, a – константы степенной аппроксимации роста организаций и публикаций соответственно.

Расчеты показывают, что точное количество публикаций имеет вид $a * t^b = 4.16e - 04 * t^{2.85}$, тогда как количество публикаций, оцененное на основе данных об организациях и среднем количестве публикаций, составляет $a * t^b = 2.78e - 04 * t^{2.82}$, что приводит к отклонению около 1 %. Это в свою очередь показывает, что микросвойства области миРНК согласуются с макросвойствами, и структура взаимодействия организаций в определенной степени определяет активность публикаций.

С учетом публикаций, также имеющих модель степенного закона, мы сравнили ведущие страны и показали, что США имели быстрый старт публикаций в этой области, достигая примерно 1000 публикаций в год в 2006 г (5 лет после первого года публикации – 2001 г.). В то время как Китай имеет более высокий общий рост, что привело к тому, что Китай стал новым лидером в 2013 г. Кроме того, большинство лидеров испытало снижение активности публикаций в последние 3–4 года.

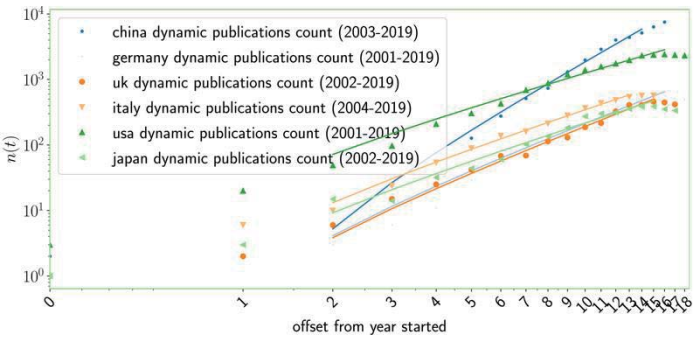


Рис. 4. Ежегодные публикации США, Китая, Германии, Италии, Великобритании, Японии в науке о миРНК со степенными приближениями. 0 – это год, когда страна опубликовала первую работу. Эти годы можно увидеть в легенде рисунка. Первые две и последние две точки были исключены из приближения. Данные каждой страны нормализованы к максимальному числу кумулятивных публикаций этой конкретной страны

Заключение: В данной работе мы показали, что граф взаимодействия научных учреждений, связанных с миРНК, следует эпидемической модели [2] и организован через эффект малого мира, а также демонстрирует степенной рост, что тоже проявляется на уровне публикаций. Основываясь на аппроксимации степенным законом, мы показали, что лидерами этой области являются США и Китай, причем последний является текущим лидером.

miRNA science as an intellectual epidemy

Firsov A.[1]*, Titov I.[2, 3]

[1] *A.P. Ershov Institute of Informatics Systems, SB RAS, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*
** artemijfirsov@mail.ru*

Key words: *k*-mer; *n*-gram; dbscan; identification; miRNA; timsort; kofer; digital library; organization co-authorship; small world; power growth; epidemic

Motivation and Aim: Digital form of scientific articles allows for automatic querying of their metadata, which can be used in estimating the statistics of the scientific field. In our work, we focus on the miRNA science field, for which we built the organization co-authorship network to provide higher-level analysis [1] of scientific community evolution compared to the author-level characteristics. In this work we show that the development of the miRNA science field is described by the epidemic model [2], that extends SIR model spreading onto transmission graph. Previously we showed that the growth of the miRNA field is governed by the small-world architecture [3], and in this work we provide the analysis of the power-law growth of publications within the area, and show the country-level comparison, outlining USA and China as the leaders.
Epidemic Knowledge Spreading: Diseases spread when infective material meets a susceptible population, consisting of infectives, susceptibles, and removals. Exposure can lead to immunity or infection, with a latency period before symptoms appear. Epidemics arise when there's a spike in cases. Goffman and Newill [4] expanded on this, discussing idea epidemics like religious or scientific concepts. Unlike desirable idea epidemics, disease outbreaks are undesirable. In scientific or lay populations, vulnerable members can catch a concept, while immune individuals may reject it or spread it unknowingly. When a significant portion of the population adopts a concept, a knowledge epidemic occurs. Such concept was further developed and applied to the transmission graphs in [2], which is discussed in the next section.

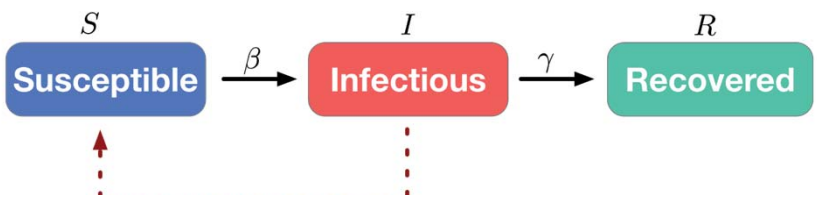


Fig. 2. Transition from Susceptible to Infectious to Recovered states in the SIR model

Dynamics of the Small-world Networks: The dissemination of knowledge can be likened to a contagion, wherein organizations are vulnerable to being “infected” by certain research fields and subsequently producing papers. These dynamics are often represented using Susceptible, Infectious, or Recovered (SIR) models [6], which demonstrate exponential growth in infections during initial stages. Nevertheless, these models do not account for the underlying graph of disease transmission.
In [2], the author models the incidence growth for problems, where the transmission graphs are known and have the small-world property [7], more detailed explanation was given in [3], and the resulting model for “infected” organizations is:

$$\rho(t) = \lambda \frac{(\lambda t)^{D-1}}{(D-1)!} e^{-(\lambda-\mu)t} \left[1 + O\left(\frac{t_0}{t}\right)\right],$$

where $\lambda/\ \mu$ – exponent parameters of infection/removal events of the SIR model, D – average shortest path length of the graph, t_0 – crossover time between power-law growth and exponent deceleration.

Methods and Algorithms: Our approach involved utilizing the PubMed digital library to compile a dataset on miRNA research. Affiliations mentioned in these papers were clarified by employing lexicographical sorting on pre-normalized and split K-Mer Boolean feature vectors. These clustered affiliations encompass references to identical organizations, facilitating the construction of a co-authorship graph to identify organizations that collaborated on publications.

Results: As shown previously [3], the miRNA graph has the small-world property, and with $D = 3.46$, and the approximated exponentiation parameter $D - 1 \approx 2.6$, it shows 5.5 % deviation from the model's predicted.

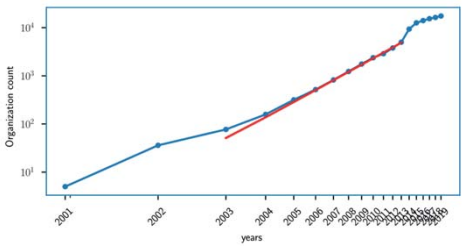


Fig. 2. Yearly organizations count, which published a paper within a certain year

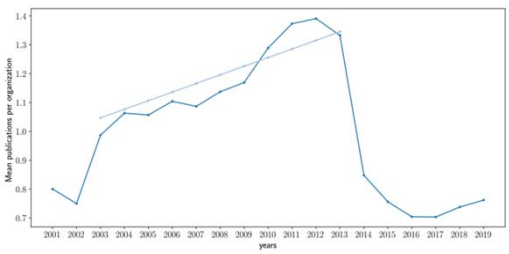


Fig. 3. Average publications per organization per year

In this work, we show that the average publications per organization follow a linear growth pattern, which can be approximated and used to compare the exact publication dynamics with estimated using the power-law model, and the dynamics of the average publications per organization:

$$a_{org} * x_{org}^{b_{org}} * (mx + l) = a * t^b,$$

where $mx + l$ – average publications per organization per year, b_{org} – organizations count exponentiation parameter, b – publications count exponentiation parameter, a_{org}, a – constants of power approximation of organizations and publications respectively.

The calculations show that the exact publications count has $a * t^b = 4.16e - 04 * t^{2.85}$, whereas publications count estimated based on the organizations and average publications data gives $a * t^b = 2.78e - 04 * t^{2.82}$, resulting in the deviation to be around 1 %, which shows that micro properties of the miRNA field are consistent with the macro properties, and the small-world structure of organization interaction defines the publication activity.

And with publications shown to also support the power-law model, we compared the leader countries using this model, showing that the USA had the rapid start of publication in this field, reaching approximately 1,000 publications per year in 2006 (5 years after the first year of publication – 2001). Whereas China has higher overall growth, which led to China becoming the new leader in 2013. Additionally, most of the leaders experienced the decline of publication activity in the late 3–4 years.

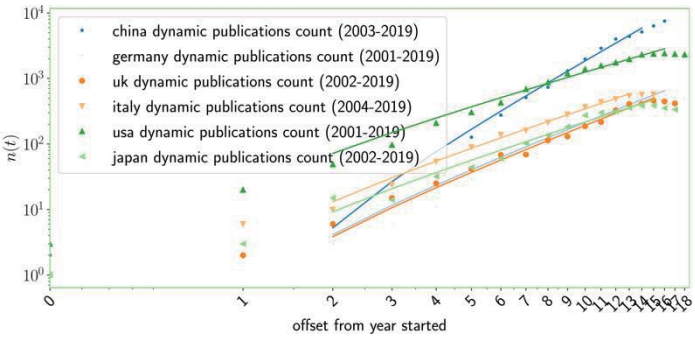


Fig. 4. Yearly publications of USA, China, Germany, Italy, UK, Japan in the miRNA science per year with power approximations. 0 is the year the country had the first publication. These years can be seen in the legend section of the figure. First two and last two points were excluded from the approximation. The data of each country was normalized to the maximal number of cumulative publications of this specific country

Conclusion: In this work we showed that the miRNA's scientific institution interaction graph follows the epidemic model [2] and is organized via the small-world effect and shows the power-law growth, which also manifests itself on the publications level. Based on the power-law approximation, we showed that the leaders of the field are USA, and China, with the latter being the current leader.

Список литературы/References

1. Leydesdorff L., Wagner C., Park H., Adams J. International collaboration in science: The global map and the network. *El Profesional de la Informacion*. 2013;22(1):87-94. doi 10.3145/epi.2013.ene.12
2. Vazquez A. Spreading dynamics on small-world networks with connectivity fluctuations and correlations. *Physical Rev*. 2006;74:056101
3. Firsov A., Titov I. Small world effect of the miRNA science field drives its growth. In: Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022). Novosibirsk: ICG SB RAS, 2022;1090-1091. doi 10.18699/BGRS/SB-2022-000
4. Goffman W. Mathematical approach to the spread of scientific ideas – the history of mast cell research. *Nature*. 1966;212(5061):449-452. doi 10.1038/212449a0
5. Ribeiro L., Rapini M., Silva L., Motta Albuquerque E. Growth patterns of the network of international collaboration in science. *Scientometrics*. 2018;114:159-179
6. Beckley R., Weatherspoon C., Alexander M. et al. Modeling epidemics with differential equations. 2013. [https://www.tnstate.edu/mathematics/mathreu/filesreu/GroupProjectSIR.pdf]
7. Humphries M.D., Gurney K. Network 'small-world-ness': A quantitative method for determining canonical network equivalence. *PLoS One*. 2008;3(4):e0002051

doi 10.18699/bgrs2024-12.3-12

От генома до оптимального сбалансированного рациона

Шлихт А.*, Краморенко Н.

Дальневосточный федеральный университет, Владивосток, Россия

** schliht@mail.ru*

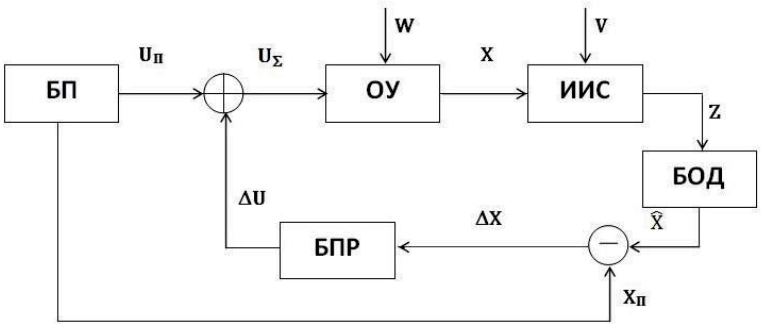
Ключевые слова: геномика; феномика; фудомика; оптимальный синтез рационов; базы данных; базы знаний; экспертные системы

Мотивация и цель: Целью исследования является создание интегрированной интеллектуальной биоинформационной системы, позволяющей проследить цепочку омиксных биохимических процессов, включая: геномику, эпигеномику, транскриптомику, экзомику, протеомику, метаболомику, реактомику, геномонтологию, геномные мутации и далее перейти к диагностике генетических заболеваний, анализу и синтезу оптимальных сбалансированных рационов. Существующие хранилища омиксных данных на базе международных порталов (NCBI, KEGG, Ensembl, UniProt, GO и др.) не обеспечивают в должной мере автоматический анализ и интерпретацию омиксных данных. Решающее значение здесь приобретают технологии искусственного интеллекта (ИИ) на основе баз данных (БД) и баз знаний (БЗ). Омик-центрированные биологические объекты и процессы стали вызовом для математики, информатики, теории управления. Учет геномного статуса индивида является важной особенностью современной теории и практики наук о жизни. Мутации в генах дают мутации в соответствующих белках. Если этот белок фермент, то происходит нарушение биохимических реакций, в которых участвует фермент, далее нарушение работы метаболических путей и физиологических процессов, в частности, происходит избыточное накопление метаболитов, которые являются причиной заболевания. Часто сбалансированное питание способно нейтрализовать последствия геномных патологий, путем создания соответствующих рационов. Здесь работают другие омики: нутригенетика, нутригеномика, нутрициология, фудомика, феномика, диетология и физиология питания в рамках модели: Генотип – Среда обитания – Фенотип. Ключевым фактором в заболеваниях, связанных с питанием, является химический состав и сбалансированность рациона.

Методы и алгоритмы: Для достижения этих целей были созданы высокоструктурированные реляционные омиксные модели БД и БЗ [1, 2], алгоритмы и программы, которые позволили формализовать биохимические и физиологические данные и знания и погрузить в системно-кибернетическую структурированную модель (СКСМ), а также в расширенную модель математического программирования, позволяющие учитывать большое разнообразие норм и ограничений в рационе [3]. Созданные модели позволяют перейти от больших количественных данных BigData, характерным для омик и химического состава продуктов, блюд и рационов, к качественным малым данным SmallData – диагнозам, рационам, удобно используемым врачами, диетологами, биологами [4, 5]. Для этого в СКСМ осуществляются анализы рационов в прямой цепи управления и оптимальный синтез дополнительных рационов в цепи

обратной связи, на основе кибернетических принципов наблюдаемости, управляемости, достижимости.

Результаты: В состав предлагаемой СКСМ [3] на базе модели пространства состояний входят подсистемы: норм питания (БП); анализа рациона (ИИС); оценки рациона (БОД); оптимальный регулятор в цепи обратной связи (БПР) (рис. 1). С векторами: X, $\hat{X}$ – соответственно векторы состояния, оценки состояния, Z – измерения, W – возмущения, V – ошибок измерения, ΔX – отклонение от норм, ΔU – дополнительное управление, Uп – плановые продукты, Xп – плановые нутриенты.



БП – блок планирования формирующий нормативные требования к управляемым объектам и процессам ($U_П, X_П$)
ОУ – объект или процесс управления с векторами управления, состояния и возмущения ($U_Σ, X, W$)
ИИС – информационно-измерительная система с векторами измерения и ошибок измерения (Z, V)
БОД – блок оценивания и диагностирования с вектором оценки состояния ($\hat{X}$)
БС – блок сравнения оценки текущего состояния и плана ($\Delta X = \hat{X} - X_П$)
БПР – регулятор в цепи обратной связи – блок принятия дополнительных решений (ΔU)

Рис. 1. Системно-кибернетическая структурированная модель

СКСМ строится по технологии глубоких экспертных систем на основе БД и БЗ с дополнительной оптимизацией на базе методов математического программирования (реализация задачи о диете). Оптимальный регулятор в цепи обратной связи определяет дефицитные нутриенты и дает рекомендации по качественному и количественному составу дополнительных продуктов.

Анализ традиционных рационов.

Применительно к задаче анализа и оптимального синтеза рациона в данном случае не учитываются возмущения (колебания химического состава продуктов), а также ошибки измерений, возникающие в процессе взвешивания и анализа химического состава продуктов. Это существенно упрощает задачу и позволяет перейти от стохастических моделей к детерминированным. Анализ химического состава традиционного рациона производится на базе системы алгебраических уравнений:

$$X = A \cdot U,$$

где U – вектор потребленных продуктов в рационе; X – вектор нутриентов;

$A = (a_{ij})$ – матрица $m \times n$, где a_{ij} – элементы матрицы, характеризующие количество i-го нутриента в j-ом продукте, m – количество анализируемых нутриентов, n – количество продуктов.

Оптимальный синтез корректирующих рационов.

Далее проводится сравнение результатов анализа с требуемыми нормами и определяется дефицит необходимых нутриентов $\Delta X = X - X_П$, т.е. вместо $\hat{X}$ в СКСМ используем просто X.

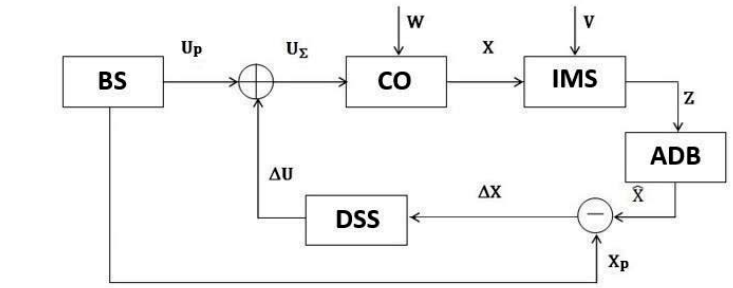
После этого решаем задачу оптимального синтеза добавления дефицитных нутриентов на основе методов линейного программирования путем минимизации линейной функции: $K = C^T \cdot \Delta U$,
где $\Delta U \geq 0$, $A \cdot \Delta U \geq \Delta X$, $C^T = (c_1, c_2, \dots, c_n)$ – вектор-строка $1 \times n$, характеризующая стоимость соответствующих продуктов.
Для решения задач анализа и оптимального сбалансированного синтеза рациона созданы БД и БЗ: норм питания, сырья, рецептур блюд, химического состава сырья и конкретных блюд, норм и ограничений в питании, суточных рационов, введены формулы питания (по А.А. Покровскому). Разработанные БД и БЗ интегрированы с алгоритмами и программами реализации симплекс метода. Для автоматического построения многомерных моделей (размерность десятки и сотни) разработаны запросы, которые вошли в состав дополнительных ограничений задачи линейного программирования и обеспечивают автоматический синтез многомерных моделей. В качестве критерия оптимизации могут быть использованы различные критерии: стоимости, минимизации калорий, веса, ограничений не только на нутриенты и их соотношения, но и на потребляемые первичные продукты.
На основе моделей баз данных и знаний, экспертных систем осуществляется: интеллектуальный анализ и интерпретация омиксных данных; формируются оптимальные сбалансированные рационы. При этом прослеживается вся цепочка причинно-следственных связей и событий в омиксных биохимических и физиологических процессах, а также синтезе оптимального сбалансированного рациона. Таким образом, мы имеем систему объяснений, характерную для экспертных систем, в отличие от интеллектуальных систем на базе обучаемых нейронных сетей, которые работают, как «черный ящик».
Выводы: Созданная биоинформационная система позволяет эффективно интегрировать многочисленные дифференциальные фундаментальные знания и внедрить в практику интеллектуальные методы анализа и интерпретации омиксных данных, автоматизировать процессы анализа и оптимального сбалансированного синтеза рационов, реализовать любую существующую диету. Для этого в автоматическом режиме создаются модели данных из созданных БД и многочисленных ограничений на сырье, первичные продукты, диеты. В случае, если не удается синтезировать рацион на базе традиционных продуктов, т.е. рацион не управляем, либо не достижим (превышение массы, избыточные калории, нутриенты, ограничения в продуктах и т.д.), то можно добавлять в модель различные пищевые добавки, витамины. Система сама определит, что и в каком количестве подходит для сбалансированности синтезируемого рациона. Достоинством разработанной системы является то, что она не ограничивает индивида строгими рамками рациона, который получается на базе оптимального синтеза, но позволяет питаться по традиционной схеме питания, но с дополнительными продуктами, получаемыми в цепи обратной связи для коррекции рациона. Обычно, это составляет значительно меньшую долю общих затрат на питание, но требует более строгого подбора сырья и готовых продуктов, а также дисциплины питания.

From the genome to an optimal balanced diet

Shlikht A.*, Kramorenko N.
Far Eastern Federal University, Vladivostok, Russia
** schliht@mail.ru*

Key words: genomics; phenomics; foodomics; optimal synthesis of diets; databases; knowledge bases; expert systems

Motivation and Aim: The goal of the study is to create an integrated intelligent bioinformatics system that allows you to trace the chain of omics biochemical processes, including: genomics, epigenomics, transcriptomics, exomics, proteomics, metabolomics, reactomics, genomontology, genomic mutations and then proceed to the diagnosis of genetic diseases, analysis and synthesis of optimal balanced diets. Existing omics data repositories based on international portals (NCBI, KEGG, Ensemble, UniProt, GO, etc.) do not adequately provide automatic analysis and interpretation of omics data. Artificial intelligence (AI) technologies based on databases (DB) and knowledge bases (KB) are becoming crucial here. Omics-centered biological objects and processes have become a challenge for mathematics, informatics and control theory. Taking into account the genomic status of an individual is an important feature of the modern theory and practice of life sciences. Mutations in genes produce mutations in the corresponding proteins. If this protein is an enzyme, then there is a disruption of the biochemical reactions in which the enzyme participates, further disruption of the metabolic pathways and physiological processes, in particular, there is an excessive accumulation of metabolites that are the cause of the disease. Often, a balanced diet is able to neutralize the effects of genomic pathologies by creating appropriate diets. Other omics work here: nutrigenetics, nutrigenomics, nutritionology, foodomics, phenomics, dietetics and nutrition physiology within the framework of the model: Genotype – Habitat – Phenotype. The key factor in nutrition-related diseases is the chemical composition and balance of the diet.
Methods and Algorithms: To achieve these goals, highly structured relational omics models of DB and KB [1, 2] were created, along with algorithms and programs. They made it possible to formalize biochemical and physiological data and knowledge and immerse them in the system-cybernetic structured model (SCSM), as well as in an extended model of mathematical programming, which allow taking into account a wide variety of norms and restrictions in the diet [3]. The created models make it possible to move from large quantitative data BigData, characteristic of omics and the chemical composition of products, dishes and diets, to high-quality SmallData – diagnoses, diets, conveniently used by doctors, nutritionists, biologists [4, 5]. For this purpose, the SCSM analyzes rations in the direct control chain and the optimal synthesis of additional rations in the feedback chain, based on the cybernetic principles of observability, controllability, and reachability.
Results: The proposed SCSM [3] based on the state space model includes subsystems: nutrition standards (BS); diet analysis (IMS); diet assessment (ADB); optimal regulator in the feedback control (DSS) (Figure 1). With vectors: X, $\hat{X}$ – vectors of state, condition assessment, respectively, Z – measurements, W – perturbations, V – measurement errors, ΔX – deviation from norms, ΔU – additional control, Up – planned products, Xp – planned nutrients.



BS – Block Scheduling – forms regulatory requirements for control objects and processes ($\mathbf{U_p, X_p}$)
CO – Control Object with control, state and perturbation vectors ($\mathbf{U_\Sigma, X, W}$)
IMS – Information-Measuring System with measurement vectors and measurement errors (**Z, V**)
ADB – Assessment and Diagnostics Block with a vector of condition assessment ($\widehat{\mathbf{X}}$)
BC – Block Comparing the assessment of the current state with the plan ($\Delta\mathbf{X} = \widehat{\mathbf{X}} - \mathbf{X_p}$)
DSS – Decision Support System – a regulator in the feedback control ($\Delta\mathbf{U}$)

Fig. 1. The System-Cybernetic Structured Model

SCSM is based on the technology of deep expert systems based on DB and KB with additional optimization based on mathematical programming methods (implementation of the diet problem). The optimal regulator in the feedback control identifies deficient nutrients and provides recommendations on the qualitative and quantitative composition of additional products.

Analysis of traditional diets.

In relation to the problem of analyzing and optimally synthesizing a diet, in this case, disturbances (fluctuations in the chemical composition of foods), as well as measurement errors that occur during weighing and analyzing the chemical composition of food products, are not considered. This greatly simplifies the task and allows you to move from stochastic models to deterministic ones. The analysis of the chemical composition of traditional diet is conducted based on a system of algebraic equations:

$X = A \cdot U$,

where U – the vector of consumed products in the diet; X – the nutrient vector;

$A = (a_{ij})$ – the matrix $m \times n$, where a_{ij} – the elements of the matrix that characterize the amount of the i-th nutrient in the j-th product, m – the number of analyzed nutrients, n – the number of products.

Optimal synthesis of corrective rations.

Next, the results of the analysis are compared with the required standards and the deficiency of necessary nutrients is determined $\Delta X = X - X_p$, i.e., instead of $\widehat{X}$ in SCSM we simply use X.

After that, we solve the problem of optimal synthesis of the addition of deficient nutrients based on linear programming methods by minimizing the linear function: $K = C^T \cdot \Delta U$, where $\Delta U \geq 0$, $A \cdot \Delta U \geq \Delta X$, $C^T = (c_1, c_2, \ldots, c_n)$ – vector-line $1 \times n$ characterizing the values of the respective products.

To solve the problems of analysis and optimal balanced synthesis of the diet, DB and KB were created: nutrition norms, raw materials, recipes of dishes, chemical composition of raw materials and specific dishes, norms and restrictions in nutrition, daily rations, nutrition formulas were introduced (according to A.A. Pokrovsky). The

developed DB and KB are integrated with algorithms and programs for implementing the simplex method. For the automatic construction of multidimensional models (tens and hundreds of dimensions), queries have been developed that are part of additional constraints of the linear programming problem and provide automatic synthesis of multidimensional models. Various criteria can be used as an optimization criterion: cost, calorie minimization, weight minimization, restrictions not only on nutrients and their ratios, but also on consumed primary products.

Based on database models and knowledge bases, as well as expert systems, intelligent analysis and interpretation of omics data is performed, and optimal balanced diets are generated. At the same time, the entire chain of cause-and-effect relationships and events in the omics biochemical and physiological processes, as well as the synthesis of an optimal balanced diet, is traced. Thus, we have a system of explanations that is typical for expert systems, in contrast to intelligent systems based on trained neural networks, which operate like a "black box".

Conclusion: The created bioinformatics system makes it possible to effectively integrate numerous differential fundamental knowledge and put into practice intelligent methods of analysis and interpretation of omics data, automate the processes of analysis and optimal balanced synthesis of diets, implement any existing diet. To do this, data models are automatically created from the databases created and numerous restrictions on raw materials, primary products and diets. If it is not possible to synthesize a diet based on traditional products, i. e. the diet is not manageable or achievable (excess weight, excess calories, nutrients, food restrictions, etc.), then various dietary supplements and vitamins can be added to the model. The system itself will determine what and in what amount is suitable for the balance of the synthesized diet. The advantage of the developed system is that it does not restrict the individual to strict dietary limits, which are obtained through optimal synthesis, but rather allows them to eat according to a traditional nutritional plan, with additional products derived from the feedback control to correct the diet. This usually accounts for a much smaller proportion of the total cost of nutrition, but requires more rigorous selection of raw materials and finished products, as well as nutritional discipline.

Список литературы/References

1. Shlikht A.G., Kramorenko N.V. Artificial intelligence in the problems of analysis and interpretation of omics human data. *BGRS\SB*-2018. 2018;78. doi 10.18699/BGRSSB-2018-054
2. Shlikht A.G., Kramorenko N.V. Deep bioinformatics expert system of analysis, modeling and interpretation of omics BigData of the human genome. MM-HPC-BBB-2018. 2018;66. doi 10.18699/MM-HPC-BBB-2018-57
3. Шлихт А.Г., Краморенко Н.В. Геном-центрированная интеллектуальная системно-кибернетическая структурированная модель в задачах оптимального синтеза рационов человека. *Актуальные вопросы биологической физики и химии*. 2020;5(2):301-304
   [Shlikht A.G., Kramorenko N.V. Genome-centered intelligent system-cybernetic structured model for optimal synthesis of human diet. *Aktualnyye voprosy biologicheskoy fiziki i khimii. = Russian Journal of Biological Physics and Chemistry.* 2020;5(2):301-304 (in Russian)]
4. Tamkovich S.N., Tutanov O.S., Serdukov D.S. et al. Protein Content of Circulating Nucleoprotein Complexes. *Adv Exp Med Biol.* 2016;924:133-136. doi 10.1007/978-3-319-42044-8_26
5. Naryzhny S., Volnitskiy A., Kopylov A., Zorina E., Kamyshinsky R., Bairamukov V., Garaeva L., Shlikht A., Shtam T. Proteome of Glioblastoma-Derived Exosomes as a Source of Biomarkers. *Biomedicines*. 2020;8(7):216. doi 10.3390/biomedicines8070216

doi 10.18699/bgrs2024-12.3-13

DeNovo sampling and sequence-derived structural descriptors based *in-silico* identification of host-pathogen protein interactions

Abbasi W.A.[1]*, Ali M.[1], Shabir A.[2], Andleeb S.[3], Bibi M.[1], Abbas S.A.[1]

[1] *Computaional Biology and Data Analysis Lab., Department of Computer Sciences & Information Technology, King Abdullah Campus, University of Azad Jammu & Kashmir, Muzaffarabad, Pakistan*

[2] *Department of Software Engineering, Capital University of Science & Technology (CUST), Islamabad, Pakistan*

[3] *Biotechnology Lab., Department of Zoology, King Abdullah Campus, University of Azad Jammu & Kashmir, Muzaffarabad, Pakistan*

** wajidarshad@ajku.edu.pk*

Key words: machine learning; protein interactions; infectious diseases; host-pathogen interactions; predictive modeling; protein sequence

Motivation and Aim: Infectious diseases pose an incessant socio-economic burden on our healthcare systems as we witnessed during the COVID-19 pandemic [1, 2]. Protein-protein interactions are a key mechanism in pathogen infections of hosts such as viral or bacterial diseases in humans [3]. Therefore, the prediction of host-pathogen protein-protein interactions (HPPPIs) is crucial for understanding the mechanisms of infectious diseases and developing effective therapies. Experimental methods for identifying these interactions are time-consuming and expensive, and there is a need for efficient computational methods to predict these interactions.

Methods and Algorithms: In this study, we propose a novel approach for predicting HPPPIs that combines sequence-derived structural features with de novo sampling for negative examples. Firstly, we extract structural features from the amino acid sequences of the host and pathogen proteins using a computational tool. These features include pseudo-amino acid compositions (PseAAC), autocorrelation features: normalized Moreau–Broto autocorrelation, Moran autocorrelation, Geary autocorrelation, composition, transition, distribution, and Sequence-order [4]. Then, we develop a machine learning model that combines these structural features with the DeNovo sampling of non-interacting proteins to predict HPIs. The DeNovo sampling technique and sequence-derived structural features increase the likelihood of identifying novel HPIs and improve the accuracy of the predictions. To evaluate the performance of our method, we apply it to a dataset of known HPIs and compare its results with those of other state-of-the-art methods. According to our findings, our method performs better than other methods in terms of accuracy and sensitivity.

Results: With leave-one-family-out (LOFO) cross-validation we have used three different classifiers (SVM, RF, and XGBoost) and three different types of feature representations (Composition, Substitution, and SDSD) to evaluate the performance of our proposed machine learning based model for host-pathogen protein interaction prediction both in terms of area under the ROC curve (AUC-ROC) and area under the PR curve (AUC-PR). For bacteria as host, our proposed model for host-pathogen protein interaction prediction achieved a maximum AUC-PR value of 76.55 % using XGB along

with sequence-derived structural feature representation. Similarly, for a virus as host, our proposed model for host-pathogen protein interaction prediction achieved a maximum AUC-PR value of 77.21 % using XGB along with sequence-derived structural feature representation. On an external validation dataset our proposed model for predicting host-pathogen interactions performed better, with AUC-PR values for viruses and bacteria of 77.77 % and 83.38 %, respectively. We have also used some benchmark datasets such as Zhou's H1N1 and Zhou's Ebola [5], Denovo Slim [6], and Barman [7] to compare our method (Proposed) with some state-of-the-art methods such as GENERALIZED [5], DOC2VEC [8], DENOVO [6], and MTT [1] by using F1 score.
Conclusion: The use of sequence-derived structural descriptors provides a more detailed representation of the proteins involved in the interaction, which is more informative than simply using the amino acid sequence. Our approach also includes negative samples generated using a de novo sampling approach, which is a valuable addition to the existing methods that rely on random negative examples.

References

1. Dong T.N., Brogden G., Gerold G., Khosla M. A multitask transfer learning framework for the prediction of virus-human protein–protein interactions. *BMC Bioinformatics*. 2021;22:572
2. Standl F., Jöckel K.-H., Brune B., Schmidt B., Stang A. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis*. 2021;21:e77
3. Acharya D., Dutta T.K. Elucidating the network features and evolutionary attributes of intra- and interspecific protein–protein interactions between human and pathogenic bacteria. *Sci Rep*. 2021;11:190
4. Cao D.-S., Xu Q.-S., Liang Y.-Z. propy: a tool to generate various modes of Chou's PseAAC. *Bioinformatics*. 2013;29:960-962
5. Zhou X., Park B., Choi D., Han K. A generalized approach to predicting protein-protein interactions between virus and host. *BMC Genomics*. 2018;19:4924
6. Eid F.-E., ElHefnawi M., Heath L.S. DeNovo: virus-host sequence-based protein-protein interaction prediction. *Bioinformatics*. 2016;32(8):1144-1150. doi 10.1093/bioinformatics/btv737
7. Barman R.K., Saha S., Das S. Prediction of Interactions between Viral and Host Proteins Using Supervised Machine Learning Methods. *PLoS One*. 2014;9:e112034
8. Yang X., Yang S., Li Q., Wuchty S., Zhang Z. Prediction of human-virus protein-protein interactions through a sequence embedding-based machine learning method. *Comput Struct Biotechnol J*. 2020;18:153-161

doi 10.18699/bgrs2024-12.3-14

iSeq: an integrated tool to fetch public sequencing data

Chao H.Y.[1], Li Z.J.[2], Chen D.J.[2], Chen M.[1]*

[1] *Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou, China*

[2] *Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, China*

* *mchen@zju.edu.cn*

Key words: iSeq; NGS; metadata; SRA; FASTQ

Motivation and Aim: High-throughput DNA sequencing technologies have revolutionized biological research [1]. Despite the significant volume of publicly available sequencing data, accessing it programmatically remains challenging. In this context, we introduce iSeq, a tool designed to streamline the retrieval of metadata and NGS data from various databases. iSeq aims to facilitate efficient data access, supporting multiple accession formats and offering functionalities such as FASTQ file merging and multi-threaded processes.

Methods and Algorithms: iSeq is a Bash script that supports the retrieval of sequencing data and metadata from GSA [2], SRA [3], ENA [4], and DDBJ [5] databases. It offers a range of features including support for multiple accession formats (Project, Study, Sample, Experiment, or Run), metadata download, direct retrieval of gzip-formatted FASTQ files, multi-threaded processes, and parallel downloads. These features enhance the efficiency and flexibility of data retrieval, catering to diverse research needs.

Results: Utilizing iSeq, we conducted a series of analyses to showcase its capabilities in retrieving public sequencing data and metadata. In one instance, we aimed to investigate the transcriptomic profiles of different tissues in a model organism. Leveraging iSeq, we effortlessly retrieved RNA-seq data from the SRA database using the corresponding accession IDs. The tool's ability to handle multiple accession formats streamlined the retrieval process, enabling us to obtain comprehensive datasets encompassing diverse tissue types. Furthermore, we explored the microbial diversity in various environmental samples by retrieving metagenomic data from the ENA database. iSeq facilitated the retrieval of metagenomic datasets associated with specific environmental conditions, allowing for in-depth analyses of microbial community structures and functional potentials. In another scenario, we sought to investigate the genetic basis of a complex trait by analyzing whole-genome sequencing data. iSeq proved instrumental in retrieving large-scale genomic datasets from the GSA database, providing access to a wealth of genetic information for downstream analyses. Moreover, iSeq's support for direct download of gzip-formatted FASTQ files expedited data retrieval, particularly for projects hosted on the GSA database. This feature proved invaluable for projects requiring rapid access to sequencing data in a compressed format, minimizing download times and storage requirements. Overall, our experiences demonstrate the versatility and efficacy of iSeq in accessing public sequencing data across multiple databases. The tool's intuitive interface, coupled with its robust functionality, empowers researchers to efficiently retrieve and utilize diverse genomic datasets for various biological inquiries.

Conclusion: iSeq presents a valuable resource for researchers seeking to access and utilize public sequencing data efficiently. By providing seamless access to multiple

databases and offering advanced features for data retrieval, iSeq empowers researchers to explore and analyze genomic datasets effectively, thereby accelerating biological research.

Funding: This study was supported by the National Natural Sciences Foundation of China (No. 32070677, 32270709, 32261133526, 32300532).

References

1. Karouia F., Peyvan K., Pohorille A. Toward biotechnology in space: High-throughput instruments for in situ biological research beyond Earth. *Biotechnol Adv*. 2017;35(7):905-932. doi 10.1016/j.biotechadv
2. Chen T., Chen X., Zhang S. et al. The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. *Genomics Proteomics Bioinf*. 2021;19(4):578-583. doi 10.1016/j.gpb.2021.08.001
3. Katz K., Shutov O., Lapoint R., Kimelman M., Brister J.R., O'Sullivan C. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res*. 2022;50(D1):D387-D390. doi 10.1093/nar/gkab1053
4. Cummins C., Ahamed A., Aslam R. et al. The European Nucleotide Archive in 2021. *Nucleic Acids Res*. 2022;50(D1):D106-D110. doi 10.1093/nar/gkab1051
5. Fukuda A., Kodama Y., Mashima J., Fujisawa T., Ogasawara O. DDBJ update: streamlining submission and access of human data. *Nucleic Acids Res*. 2021;49(D1):D71-D75. doi 10.1093/nar/gkaa982

doi 10.18699/bgrs2024-12.3-16

Programming methodology for trusted artificial intelligence

Goncharov S.[1†]*, Nechesov A.[1†]**, Sviridenko D.[2]***

[1] *S.L. Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia*

[2] *Institute of Philosophy and Law, SB RAS, Novosibirsk, Russia*

** s.s.goncharov@math.nsc.ru; ** nechesov@math.nsc.ru; ***dsviridenko47@gmail.com*

Key words: trusted AI; programming methodology; learning theory

The task of building reliable, transparent and explainable artificial intelligence attracts the attention of all the world's researchers. Today it becomes clear that all advanced large language models do not satisfy even the basic requirements put forward for trusted intelligent systems.

The report will touch upon the development of trusted artificial intelligence algorithms [1]. A programming methodology in Turing complete languages will be presented, which is based on the task approach and the concept of semantic programming [2–5].

Our task is to make the thinking process of intelligent machines transparent and human-understandable. For these purposes, it is also necessary to build a coherent and high-quality concept of learning theory [6]. This concept should allow, based on logical rules and basic statements, to show the user the final result produced by the intelligent system. For these purposes, a new logical-probabilistic method will be presented, which allows us to close all these gaps.

Funding: The work was performed within the state task of the S.L. Sobolev Institute of Mathematics (project No. FWNF-2022-0011).

References

1. Vityaev E.E., Goncharov S.S., Gumirov V.S., Mantsivoda A.V., Nechesov A.V., Sviridenko D.I. Task approach: on the way to trusting artificial intelligence. In: World Congress Systems Theory, Algebraic Biology, Artificial Intelligence: Mathematical Foundations and Applications. Selected Works. 2023;171-243. doi 10.18699/sblai2023-41
2. Goncharov S., Nechesov A. Polynomial Analogue of Gandy's Fixed Point Theorem. *Mathematics*. 2021;9:2102. doi 10.3390/math9172102
3. Goncharov S., Sviridenko D. Logical Language of Description of Polynomial Computing. *Dokl Math*. 2019;99:121-124. doi 10.1134/S1064562419020030
4. Goncharov S., Nechesov A. Solution of the Problem P = L. *Mathematics*. 2022;10:113. doi 10.3390/math10010113
5. Goncharov S., Nechesov A. Semantic programming for AI and Robotics. In: IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation. 2022;810-815. doi 10.1109/SIBIRCON56155.2022.10017077
6. Goncharov S., Nechesov A. AI-Driven Digital Twins for Smart Cities. *Eng Proc*. 2023;58:94. doi 10.3390/ecsa-10-16223

doi 10.18699/bgrs2024-12.3-17

Enhancing biomedical knowledge discovery through hybrid text-mining and graph neural network approaches in ANDSystem

Ivanisenko T.V.[1, 2]*, Demenkov P.S.[1, 2, 3], Kolchanov N.A.[1, 2, 3], Ivanisenko V.A.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *itv@bionet.nsc.ru*

Key words: text-mining; graph neural networks; large language models; ANDSystem; Associative Gene Networks

Motivation and Aim: The exponential growth of biomedical literature complicates the extraction and understanding of complex biological systems. Traditional text-mining methods, while essential, often yield fragmented knowledge that may overlook implicit connections within biological data. This work introduces a novel hybrid approach integrated into the new version of ANDSystem [1], combining classical text-mining techniques with advanced graph neural networks (GNNs) and fine-tuned large language models (LLMs) to enhance the extraction and interpretation of biomedical information.
Methods and Algorithms: The updated ANDSystem incorporates a hybrid methodology that synergizes traditional text-mining with modern machine learning technologies, specifically graph neural networks (GNNs) and large language models (LLMs), to enhance the reconstruction and interpretation of biomedical knowledge graphs. The pipeline includes the following steps:
Named Entity Recognition: Molecular-biological entities are identified in unstructured scientific texts using ANDSystem's ontology dictionaries.
Entity Verification: The system verifies the identified names against their corresponding entity types using the context from the article, with false recognitions filtered out by the fine-tuned PubMedBERT model.
Relation Extraction: Interactions between entities are automatically established using ANDSystem's rule sets and templates, which are enhanced by information extracted from external databases.
Computation of co-occurrences: The system calculates co-occurrence values for entity pairs using the ANDDigest module to gauge the likelihood of interactions.
Graph Neural Network Training: A graph neural network is trained to represent vertices of the knowledge graph vectorially using the GraphSAGE algorithm [2].
Link Prediction Model Training: A binary classification model is trained to predict interactions between vertex pairs based on their vector representations and co-occurrence values.
Graph-based Interaction Prediction: The trained classifier is applied to predict new interactions between graph vertices.
Contextual Interaction Prediction: A fine-tuned generative LLM is used to predict interactions based on contextual cues and to provide explanations for these predictions.

The system searches for literature mentioning the interacting pairs using ANDDigest [3] and applies the fine-tuned model to detect and verify information about the connections.
Results: The application of the developed hybrid approach within ANDSystem demonstrated improved accuracy and comprehensiveness in identifying and predicting biologically relevant interactions. Our results indicate significant enhancements in the system's ability to not only identify known interactions but also predict novel ones based on contextual understanding derived from LLMs and GNNs.
Conclusion: The integrated approach of ANDSystem facilitates a more comprehensive understanding of complex biological networks. This methodology not only increases the reliability of the reconstructed networks but also provides a platform for detailed biological interpretation, thus supporting advanced research in biomedical sciences.
Funding: The study is supported by the budget project, FWNR-2022-0020.

References

1. Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(1):34. doi 10.1186/s12859-018-2567-6
2. Hamilton W., Ying Z., Leskovec J. Inductive representation learning on large graphs. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. 2017
3. Ivanisenko T.V., Demenkov P.S., Kolchanov N.A., Ivanisenko V.A. The New Version of the ANDDigest Tool with Improved AI-Based Short Names Recognition. *Int J Mol Sci*. 2022;23(23):14934. doi 10.3390/ijms232314934

doi 10.18699/bgrs2024-12.3-18

SmartCrop cognitive platform: modules for automatic knowledge extraction from scientific publications, patents and factual databases for reconstruction and analysis of gene networks of stress response in rice and wheat

Ivanisenko V.A.[1, 2]*, Demenkov P.S.[1, 2], Ivanisenko T.V.[1, 2], Antropova E.A.[1], Volyanskaya A.R.[1, 2], Zubairova U.S.[1, 2], Makarova A.A.[1, 2], Orlov Y.L.[1, 3, 4], Chen M.[5]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*
[4] *Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia*
[5] *College of Life Sciences, First Affiliated Hospital of Medical School, Zhejiang University, Hangzhou, China*
* *salix@bionet.nsc.ru*

Key words: cognitive platform; gene networks; stress response; rice; wheat

Motivation and Aim: Rice (*Oryza sativa* L.) and wheat (*Triticum aestivum* L.) have high food, technical, and feed value and are among the most important agricultural crops. Almost half of the world's population uses rice as a staple food. This product is most widespread in Asia. Wheat is the staple food for 35 % of the world's population and is the most important agricultural crop for Russia. Such widespread use of rice and wheat as food products determines their central role in ensuring global food security. However, extreme environmental impacts, including those caused by adverse climate change, plant diseases and insect pests lead to significant crop losses. The study of molecular genetic mechanisms of plant resistance to adverse biotic and abiotic factors (high or low temperature, drought, salinity, soil pollution with metals, diseases, the action of pathogens and pests, etc.) requires analysis of large multi-omics experimental data and reconstruction of gene networks, including complex signaling, regulatory, transport and metabolic pathways.

To create the cognitive software and information platform "Smart Crop", providing a full cycle of knowledge engineering in the field of plant biology necessary for solving problems of studying molecular genetic mechanisms of the genotype-phenotype-environment relationship for agriculturally valuable rice and wheat crops, the efforts of Chinese and Russian groups were combined. The Chinese group, led by Prof. Ming Chen, possesses advanced data mining information technologies for analyzing plant omics data [1–5]. The Russian team from the ICG SB RAS is a leader in the field of automatic knowledge extraction from scientific publication texts and reconstruction of gene networks [6, 7].

The platform is focused on solving such substantive tasks as interpretation of the results of omics data (establishing a connection between gene sets and biological processes, phenotypic traits, etc.); reconstruction of gene networks describing the relationship between molecular genetic objects and objects corresponding to the concepts of breeding, phenomics and seed production, phytopathology, diagnostics, protection

means, etc.; identification of regulatory and signaling pathways of plant response to specific growth conditions, biotic and abiotic stresses (high or low temperature, drought, salinity, soil contamination with metals, response to fertilizers, the effect of hormones, elicitors, etc.); prediction of candidate genes for genotyping; search for markers for marker-assisted selection; targets for directed impact of substances (including external factors) on plants to solve problems of early/uniform germination, better vegetative growth, efficient absorption of nutrients, etc.

The aim of this work is to present the results on the development of the "Smart Crop" platform module, which performs automatic analysis of texts of scientific publications, patents and factual databases and reconstruction on this basis of gene networks of rice and wheat response to stress.

Methods and Algorithms: Automatic analysis of scientific publications and patent texts was carried out using an integrated approach that includes an ontological description of the subject area and a set of methods for extracting knowledge from texts based on semantic-linguistic patterns, large linguistic neural network models, and graph neural networks. The development of semantic-linguistic patterns was carried out by adapting the patterns of the ANDSystem software and information system developed by us earlier [6]. The knowledge extracted from the texts was represented as a knowledge graph (associative gene networks) describing the relationships between the entities of the ontology. The technology was configured for the subject areas of breeding, phenomics and seed production, phytopathology, diagnostics, protection tools, etc. of wheat and rice.

Results: A knowledge extraction unit of the "Smart Crop" cognitive system has been developed, which includes modules for domain-specific ontology, information extraction from scientific texts and databases, and gene network reconstruction. Dictionaries for various entities related to rice and wheat have been created within the ontology. The initial filling of the "Smart Crop" knowledge base was performed using automatic analysis of scientific publications from 2023, utilizing semantic-linguistic patterns and information from factual databases. The analysis was conducted on a subset of articles (33,000 for rice and 29,000 for wheat) to identify errors and refine dictionaries and templates. The templates extracted over 200,000 interactions for wheat and 45,000 for rice, with "Association" and "Regulation" interactions being the most prevalent. Additionally, around 220,000 interactions were extracted from factual databases for both crops. The extracted interactions demonstrated good quality in terms of completeness and accuracy.

Conclusion: In the future, it is planned to integrate the knowledge extraction unit developed by the Russian group and the multi-omics data analysis unit developed by the Chinese group within the framework of the "Smart Crop" cognitive platform, which will provide full-featured capabilities of the platform for solving problems of knowledge engineering in the field of plant biology necessary for studying the molecular genetic mechanisms of the relationship between the genotype and phenotype of agriculturally valuable rice and wheat crops, taking into account environmental factors.

Funding: The work of IVA, DPS, ITV, AEA, VAR, ZUS and OYL was supported by the Russian-Chinese grant from the Russian Science Foundation No. 23-44-00030. The work of MCh was supported by National Natural Science Foundation of China (32261133526).

References

1. Chen D., Chen M., Altman T., Klukas C. Bridging Genomics and Phenomics. In: Approaches in Integrative Bioinformatics. 2014;299-333

2. Chen M., Liu L., Mei Q., Zhang Z. Ricenet: Genome-scale multi-level network reconstruction of rice. *J Comp Sci Syst Biol*. 2013;6:4
3. Liu L., Jiang L., Chen M. The Organelle-focused Proteomes and Interactomes in Rice. *Curr Protein Peptide Sci*. 2014;15(6):583-590
4. Liu L., Mei Q., Yu Z., Sun T., Zhang Z., Chen M. An Integrative Bioinformatics Framework for Genome-scale Multiple Level Network Reconstruction of Rice. *J Integr Bioinf*. 2013;10(2):223
5. Rahaman M.M., Chen D., Gillani Z., Klukas C., Chen M. Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front Plant Sci*. 2015;6:619
6. Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: An Associative Network Discovery System for Automated Literature Mining in the Field of Biology. *BMC Syst Biol*. 2015;9(Suppl. 2):S2
7. Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl. 1):H.34

doi 10.18699/bgrs2024-12.3-19

Inferring posterior weight distribution in hierarchical mixture models

Meshcheryakov G.[1]*, Kovanova A.V.[2]
[1] *Institute of Protein Research, RAS, Puschino, Russia*
[2] *MediaLine Group, St. Petersburg, Russia*
* *georgy@vega.protres.ru*

Key words: genotype phasing; mixture; special functions; dynamic programming; hierarchy

Motivation and Aim: Mixture models are used to model heterogenous data that is formed by sampling from multiple known distributions. The exact proportion of the data belonging to one each of the specified distributions is usually unknown and is a parameter to be estimated. In Bayesian treatment of the problem, it is assumed that mixture parameters follow a prior distribution, typically the Beta (or Dirichlet in a multivariate case) distribution. Then, posterior weight distribution is trivially inferred for each data point with a maximum-a-posteriori value usually being interpreted as a probability of a data point belonging to a particular mixture component. However, in some applications it is known that subsets of data points jointly originate from the same mixture component. Then, the posterior distribution attains a seemingly intractable form that is tough to compute. For instance, this problem arises when dealing with unphased data in high-throughput sequencing experiments. There, multiple samples/observations of the same SNP have the same unknown phase [1, 2]. The number of SNPs can be in tens of thousands and the average number of observations per SNP can be as high as a hundred, making applying stochastic algorithms such as MCMC infeasible. Here, we infer an analytical form for the posterior weight distribution of a mixture model and provide an algorithm to compute it in $O(n^2)$, making its computation possible in most practical cases.

Methods and Algorithms: Assume that we are dealing with a stratified dataset X of size N and p strata/groups $\{X^{(i)}\}^n_{i=1}$. For each strata it it is known that a data sample x from i-th strata originated either from F or G distributions with a probability of the former being w_i. That's it, the PDF of each sample being $h(x|w_i) = w_i f(x) + (1 - w_i)g(x)$. We assume that $w_i \sim Beta(\alpha, \beta)$ and we are interested in the posterior distribution $z(w_i|X^{(i)}) = \frac{L(X^{(i)}|w_i)Beta_{w_i}(\alpha,\beta)}{L(X^{(i)})}$, where $L(X^{(i)}|w_i)$ is the conditional likelihood and $L(X^{(i)})$ is the marginal likelihood. The latter is troublesome to compute when $n > 1$:

$$L(X^{(i)}) = \int_0^1 \prod L(X^{(i)}|w_i)Beta_{w_i}(\alpha,\beta)dw_i \propto \int_0^1 \sum_{0_n<k<1_n} (\frac{1-w}{w})^{|k|}t^k \propto S_t(n,\alpha,\beta),$$

where we used the multi-index notation for the multidimensional sum is a vector of ratios $\frac{f(x_i)}{g(x_i)}$, $S_t(n,\alpha,\beta) = \sum_{0_n<k<1_n} t^k\Gamma(n - |k| + \alpha)\Gamma(|k| + \beta)$. Straightforward

computation of S_t requires computation of 2^n terms. Fortunately, the S_t has properties that alleviate this problem.

Theorem 1. The following equation holds:

$$S_t(n, \alpha, \beta) = S_t(n-1, \alpha+1, \beta) + t_p * S_t(n-1, \alpha, \beta+1).$$

Theorem 2. The following equation holds:

$$S_t(n, \alpha+1, \beta) = (n+\alpha+\beta)S_t(n, \alpha, \beta) - S(n, \alpha, \beta+1).$$

Theorem 3. $\{S_t(n, \alpha, \beta+i)\}^{n+1}{}_{i=0}$ are linearly dependent, i. e. there exist such $\{C_i\}^{n+1}{}_{i=0}, C_i \neq 0$, that

$\sum_{i=0}^{n+1} C_i S(n, \alpha, \beta+i) = 0.$

Theorem 4. $\{C_i\}^{n+1}{}_{i=0}$ attain the form

$$C_i = (-1)^{n-i+1} \beta^{(n+1-i)} \frac{n+1}{i},$$

where $\beta^{(n+1-i)}$ is a raising factorial [3].

Algorithm of complexity $O(n^3)$ follows from the first theorem, whereas the second one simplifies it down to $O(n^2)$. Theorems 3 and 4 are further used to decrease the number of computations twofold, see Figure 1.

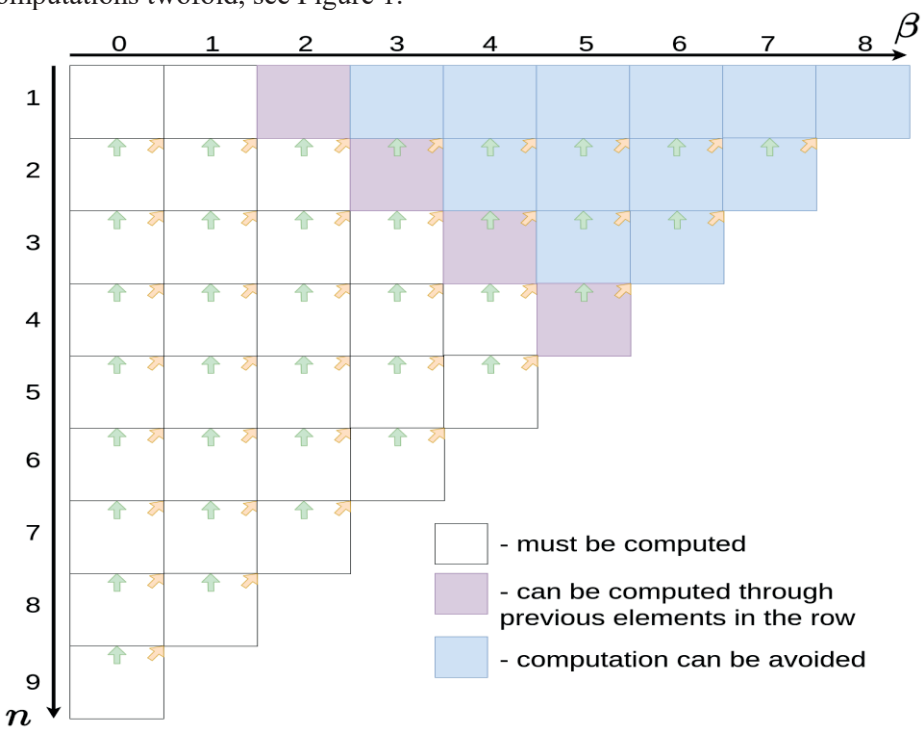


Fig. 1. Dynamic programming matrix of the algorithm that follows from theorems 1–4

References

1. Meshcheryakov G. et al. Beta negative binomial mixture model facilitates identification of allele-specific gene regulation in high-throughput sequencing data. In: Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022). Novosibirsk, 2022;1107. doi 10.18699/SBB-2022-664
2. Meshcheryakov G. et al. MIXALIME: MIXture models for ALlelic IMbalance Estimation in high-throughput sequencing data. *arXiv*. 2023. doi 10.48550/arXiv.2306.08287
3. Abramowitz M., Stegun I. (Eds.). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. NBS, 1964

doi 10.18699/bgrs2024-12.3-20

Toward improved calling of allele-specific events in high-throughput sequencing data with a mixture of compound Dirichlet negative multinomial distributions

Meshcheryakov G.*, Kulakovskiy I.V.
Institute of Protein Research, RAS, Puschino, Russia
** georgy@vega.protres.ru*

Key words: allelic imbalance; multinomial distribution; compound models; mixture model

Motivation and Aim: Heterozygous sites in gene regulatory regions may have allele-dependent chromatin states or protein affinity. High-throughput sequencing efficiently reveals nucleotide variants at homologous chromosomes of the human genome. The allelic imbalance of read counts supporting alternating alleles at individual sites reveals functional consequences of nucleotide substitutions in regulatory regions. Thus, it is possible to study the allele-specific gene regulation using the data from assays focused on genome regulatory regions, such as ChIP-Seq for transcription factor binding, or DNase- and ATAC-Seq for chromatin accessibility. Quantitative evaluation of the allelic imbalance requires a proper statistical framework accounting for noise, technical variability, and biases in the underlying read count distributions. Here we present a novel approach utilizing joint bivariate Dirichlet negative multinomial compound distributions to improve accuracy and better reflect the underlying nature of high-throughput sequencing experiments.

Methods and Algorithms: Statistical scoring of the allelic imbalance faces three challenges. First, the relative allelic copy numbers, the background allelic dosage, usually remains unknown. This challenge is solved by BABACHI [1] which performs unsupervised estimation of the relative background allelic dosage (BAD) directly from variant calls. At known BAD, the expected frequency of allelic counts is $f_1 = \frac{BAD}{BAD+1}$ and $f_2 = 1 - f_1 = \frac{1}{BAD+1}$. Then, it is reasonable to model counts distribution as a mixture of 2 distributions with frequencies f_1 and f_2.

The second challenge is to estimate whether the observed imbalance at individual variants significantly deviates from the expected imbalance arising from BAD. The straightforward approach is to use the binomial or the negative binomial distribution and estimate the success probability and the location shift parameters. Yet, those approaches fail to model the observed distributions of allelic read counts for particular types of experimental data, such as mapping accessible chromatin with ATAC-Seq or profiling transcriptional activity of regulatory elements with CAGE. In those cases, the (negative-) binomial model fails to approximate the empirical distribution adequately. The problem can be alleviated by assuming that frequencies follow the beta distribution, resulting in the beta-binomial or the beta negative binomial models [2].

The third challenge is to ensure robustness under model misspecifications. In practice, model assumptions might deviate from the true data-generating distribution. For instance, the joint datasets could include multiple sequencing batches or different cell types, each with its unique distribution parameters. Even if it is possible to specify the

model-generating process and all sources of noise, the analytical formulae for such models become intractable.

In our previous work [2, 3], we did not specify the correct model generating distribution and resorted to using the pseudolikelihood framework. There, we assumed conditional distributions $x_{|y}$ and $y_{|x}$ to be known (where (x, y) is a pair of reference and alternative allele read counts) and to follow the negative binomial or beta negative binomial distribution. Yet, estimating the model parameters from conditional distributions introduces biases into resulting estimates. This pitfall affects other common approaches, including those based on the beta-binomial distribution.

To see this, let the distribution of x given coverage $n = x + y$ follow the beta-binomial law with the success rate p and the concentration parameter κ. The log-likelihood function for the joint distribution of (x, y) is:

$$l(x, n|p, \kappa) = l(x|n, p, \kappa) + l(n|p, \kappa).$$

Maximizing it with respect to p and κ is equivalent to maximizing the conditional log-likelihood $l(x|n, p, \kappa)$ alone only if the latter marginal term doesn't depend on any of the 2 parameters. In this work we show, that that marginal likelihood, in fact, depends on the parameter κ, making its estimates obtained via maximizing the conditional log-likelihood unreliable.

We generalize the existing models under the vein of the joint bivariate distribution, based on the Dirichlet negative multinomial distribution [3] (DNM):

$$(x, y) \sim \omega\, DNM(r_0, \alpha_0, \alpha_1, \alpha_2) + (1 - \omega) DNM(r_0, DNM(r_0, \alpha_0, \alpha_2, \alpha_1)), \quad (1)$$

where α_i are parameterized as $\alpha_0 = \frac{1-p_0}{p_0}\kappa, \alpha_1 = p\,\kappa, \alpha_2 = (1-p)\kappa,$ with p_0, r_0controlling the coverage distribution, p being a parameter controlling the mapping bias ($p = f_1 = \frac{BAD}{BAD+1}$ if there is no bias), and κ being an overdispersion parameter. When dealing with data of $BAD = 1$, the mixture can be omitted by setting $w = 1$. The model can further accommodate a greater degree of overdispersion by assuming that $r_0 \sim ZTBinom(\mu, \sigma)$, where $ZTBinom$ is a zero-truncated Binomial distribution parameterized in terms of mean value μ and variance σ. It can be shown that it is possible to marginalize over r_0 and obtain an analytical form of the compound distribution [3, 4]. The parameters $p_0, p_1, p_2, r_0, \kappa$ and μ_0, σ (when dealing with the compound model) are estimated using the Local Maximum Likelihood (LML) method. Using LML instead of ML allows for capturing coverage-specific effects in the data. JAX automatic differentiation library to obtain analytical gradients of the likelihood objective function and scipy's SLSQP solver to obtain local minima.

After estimation of the model parameters, we compute p-values to identify ASEs using conditional distributions:

$$y_x \sim \omega BetaNB(r, p\kappa, (1-p)\kappa) + (1 - \omega) BetaNB(r, (1-p)\kappa, \mu\kappa), \quad (2)$$

where $r = r_0 + x$. Note that as $\kappa \to \infty$, $BetaNB$ converges to NB. Also note that alternatively, we could've conditioned on the coverage $n = x + y$ instead, obtaining a mixture of beta-binomial models instead.

If we use the compound model, then instead of $BetaNB$ we use $MCBNB$ (Marginalized Compound Beta Negative Binomial distribution) with the probability mass function

$$f(y|x, p, \kappa, r_0) \propto {}_3F_2(\kappa p + 1, y + 1, r - 1;\ 2, (1-p)\kappa - r;\ 1 - p), \quad (3)$$

where $r = r_0 + x$ and ${}_3F_2$ is a generalized hypergeometric function [5].

Conclusion: The proposed model is freely available as a part of the updated **MIXALIME** software at the Python PyPi package repository:

pip install mixalime

The detailed information on the software and documentation is available at github.com/autosome-ru/mixalime and at the website mixalime.georgy.top. The proposed method will facilitate improved reliability for identification of allele-specific regulatory events in a wide repertoire of sequencing assays, including CAGE-Seq and ATAC-Seq.

Funding: This study was supported by RSF 20-74-10075.

References

1. Abramov S. et al. Landscape of allele-specific transcription factor binding in the human genome. *Nat Commun*. 2021;12:2751. doi 10.1038/s41467-021-23007-0
2. Meshcheryakov G. et al. Beta negative binomial mixture model facilitates identification of allele-specific gene regulation in high-throughput sequencing data. In: Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022). Novosibirsk, 2022;1107. doi 10.18699/SBB-2022-664
3. Meshcheryakov G. et al. MIXALIME: MIXture models for ALlelic IMbalance Estimation in high-throughput sequencing data. *arXiv*. 2023. doi 10.48550/arXiv.2306.08287
4. Meshcheryakov G. et al. Improved identification of allele-specific gene regulation in high-throughput sequencing data with the marginalized compound negative binomial distribution. Systems Biology and Bioinformatics (SBB-2023). Novosibirsk, 2023;44. doi 10.18699/SBB-2023-44
5. Abramowitz M., Stegun I. (Eds.). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. NBS, 1964

doi 10.18699/bgrs2024-12.3-21

Mutation in of palindromic sequences of coronavirus genome

Mitić N.S.[1]*, Kapunac S.[1], Beljanski M.V.[1], Dergilev A.I.[2], Orlov Y.L.[2, 3]**

[1] *University of Belgrade, Belgrade, Serbia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

** nenad.mitic@matf.bg.ac.rs, ** orlov@bionet.nsc.ru*

Key words: computer genomics, microbiology, coronavirus, mathematical modelling, mutation rate

Motivation and Aim: The RNA genome of SARS-CoV-2 was shown to be organized into structural and functional blocks of RNA information that are demarcated by short RNA breakpoint sequences that promote recombination at specific non-random locations within the viral genome consisting of short repetitive sequences, namely palindromes. Palindromic sequences are involved in the formation of RNA secondary structures. They can be locations recognized by RNA-binding proteins as well as places of RNA recombination [1]. We analyzed SARS-COV-2 genomes with particular attention to mutations within palindromic sequences.

Methods and Algorithms: A dataset of 423.425 complete isolate nucleotide sequences was extracted from https://www.ncbi.nlm.nih.gov/sars-cov-2. After the cleanup process, 347.962 isolates with 123.667 unique (related nucleotide sequences) with 226.624 corresponding unique protein (nucleotide) coding sequences and 141.926 unique protein (AA) sequences remain. The consistency of the two sequences was checked using the standard genetic code table (transl_table 1). Each sequence was annotated with a World Health Organization (WHO) SARS-CoV-2 annotation. Each nucleotide sequence was individually aligned to the reference sequence SARS-COV -2 (NC_045512.2) using the MAFFT alignment program [2]. The StatRepeats program [3] was used to determine all palindromes with a minimum length of 8. A total of 801.935.394 palindromes were determined. Among them, 785.854.841 repeats were identical with their pair in reference sequence NC_045512.2. Other (16.080.553) palindromes have some mutations related to reference sequence. The analysis of the number of palindrome occurrences was performed in 5 time intervals of 4 months (pandemic time from 31.12.2019 to 25.08.2021). The average number of palindromes per isolate shows a constant increase, respectively by intervals: 1.92, 3.51, 9.31, 14.84, and 20.66.

Results: The pipeline for the analysis of viral genome in relation to mutation rate is presented. We analyzed mutations in all 12 types of ORFs present in the set of extracted sequences (ORF1a polyprotein, ORF1ab polyprotein, surface glycoprotein, ORF3a protein, envelope protein, membrane glycoprotein, ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, nucleocapsid phosphoprotein, ORF10 protein). Among them, normalized on average protein length, after ORF1a and ORF1ab,the surface glycoprotein (S-protein) has the highest number of repeats, on average 4.65 palindromes with a length of > =8. The highest number of palindromes is located around positions 22.000 (left part) and 24.300 (right part), counting the positions with respect to the beginning of the isolates. For the total number of mutations, almost 78% resulted in amino-acid changes in the corresponding proteins.

Conclusion: Available tools and the problems of mutation rate dependence on sequence context will be discussed. We discussed this work at VII Congress on Biophysics [4]. The complexity measures and the entropy estimates of genome segments help to reveal mutated profile of the sequences [5]. We plan to perform a detailed analysis of mutations of palindromic sequences according to the SARS-CoV-2 WHO variant classification and also their influence on amino-acid changes and occurring RNA secondary structures or locations recognized by RNA binding proteins.

Acknowledgments: The authors are grateful to the colleagues from Faculty of Mathematics, Belgrade University for science discussion.

References

1. Gallaher W.R. A palindromic RNA sequence as a common breakpoint contributor to copy-choice recombination in SARS-COV-2. *Arch Virol*. 2020;165:2341-2348. doi 10.1007/s00705-020-04750-z
2. Katoh K., Rozewicki J., Yamada K.D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2019;20(4):1160-1166. doi 10.1093/bib/bbx108
3. Jelovic A., Mitic N., Eshafah S., Beljanski M. Finding Statistically Significant Repeats in Nucleic Acids and Proteins. *J Comput Biol.* 2018;25:375-387. doi 10.1089/cmb.2017.0046
4. Anashkina A.A., Rubin A.B., Gudimchuk N.B., Vanin A.F., Tsygankov A.A., Orlov Y.L. VII Congress of Russian Biophysicists-2023. *Biophys Rev*. 2023;15(5):801-805. doi 10.1007/s12551-023-01164-4
5. Orlov Y.L., Orlova N.G. Bioinformatics tools for the sequence complexity estimates. *Biophys Rev*. 2023;15(5):1367-1378. doi 10.1007/s12551-023-01140-y

doi 10.18699/bgrs2024-12.3-22

Millions of SARS-CoV-2 genomes in the RAM of a regular PC: fast and efficient analysis of evolutionary changes

Palyanov A.Yu.[1, 2, 3]*, Palyanova N.V.[2]

[1] *A.P. Ershov Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia*

[2] *Research Institute of Virology, Federal Research Center for Fundamental and Translational Medicine, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *palyanov@iis.nsk.su*

Key words: coronavirus; SARS-CoV-2; genome; variants; evolution; analysis; software

Motivation and Aim: In addition to the enormous harm caused to humanity, the SARS-CoV-2 pandemic has also left us, thanks to the efforts of virologists and bioinformaticians, an unprecedented amount of information about how the genome of the virus changed day by day on a planetary scale for more than four years. Analysis of these data made it possible to learn a lot both specifically about SARS-CoV-2 and about respiratory viruses in general. As of May 3, 2024, there are more than $16.7 \cdot 10^6$ genetic sequences of this virus in the GISAID database, and more than $8.6 \cdot 10^6$ in NCBI Genbank. For comparison, the influenza virus is represented in the GISAID database by $5.22 \cdot 10^5$ variants (collected since 1905). Typically, to process such volumes of data, they are read directly from the HDD or SSD, but what if we could place all available SARS-CoV-2 genomes into the RAM of a regular PC and thus be able to access them and perform various analytical calculations at a much higher speed? This approach allowed us to analyze the evolution of SARS-CoV-2 throughout the pandemic.

Methods and Algorithms: One copy of the SARS-CoV-2 genome, a single-stranded RNA(+) virus, has a size of about 29900 nucleotides, and accordingly, it occupies about 29.9 kb in memory. For 16.7 million genomes, 500 GB will be required, while the typical memory capacity of a modern desktop PC is 32 – 64 GB. If alignments are calculated for all genomic variants, considering the initial SARS-CoV-2 genome as a reference sequence, then each genome can be represented as a data structure that describes the differences between considered genomic variants and the reference genome without data loss. These differences consist of a list of point mutations (a position and a nucleotide at that position after replacement), deletions (a list of intervals within which nucleotides were removed from the genome) and insertions (a list of positions where insertions occurred, and fragments of nucleotide sequences that were inserted). At each position of the aligned genome there can be either the same nucleotide as in the reference genome, or a different one (appearing as a result of mutation or insertion), or its absence (as a result of deletion), called gap ("–"). Each position that differs from the reference one increases the editing distance (or the distance between sequences) by one. Even for the variants with maximal distance from the reference genome in most cases it does not exceed 200. This representation of the data allows for compression of 1500 times relative to its original size, i. e. only 330 MB of memory. In addition, calculation of editing distance between any two sequences doesn't require comparison of all positions, only

those which differ from the reference, which provides another advantage – makes it work much faster.

Results: The implementation of the proposed approach was included in the new version (v.1.1) of the ParSeq software package previously developed by our team [1]. It now allows to read multiple SARS-CoV-2 genomes from a given set of FASTA files and align them relative to the reference sequence (using locally installed NextAlign console application [2]), convert alignment results to compact structures containing information about all the differences between a particular genome and the reference one, and also contains an algorithm for quickly and efficiently comparing a pair of any two genomes with each other. Among the mutations, deletions and insertions that determine the differences between the first and second genomes from the reference one, there can be both different and identical ones; when performing calculations, different ones are taken into account, but identical ones, respectively, are not.

Using ParSeq1.1 together with the complete dataset of SARS-CoV-2 genomes from Genbank [3], we obtained the following results:

a) Of the total number of genomes, only 21.3% do not contain unidentified nucleotides ("N") in at least one of the positions. The proportion of genomes with only one "N" is less than 1.2%, two "N"s – 0.95%, three "N"s – 0.85%, etc. (distribution of the number of genomes depending on the number "N" in them was obtained, Fig. 1).

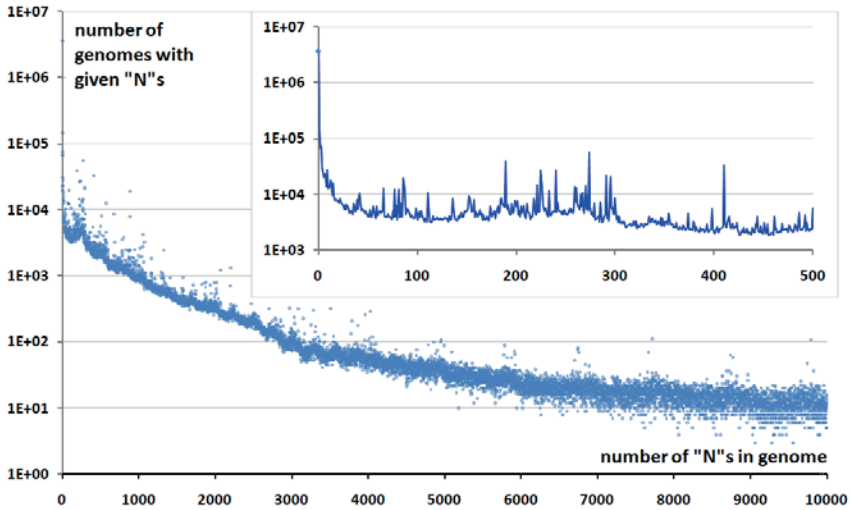


Fig. 1. Distribution of the number of sequences versus the number of "N"s in them

However, if we consider only the coding part of the genome, excluding the 3' UTR (265 nt) and 5' UTR (229 nt), then the percentage of genomes without "N" will increase to 43.3, and this is the dataset that we use in further calculations (their number is about 3.7 million).

b) The number of identical genomes (those in which the nucleotide sequence is completely identical, but the metadata may differ) was calculated. As it turns out, 54.7% of the 3.7 million genome variants are not unique – they can be represented as 459102 clusters, each composed of identical sequences, ranging in size from 2 to 12824. The distribution of cluster sizes depending on their number in the list sorted by size was calculated.

c) In Fig.e 2, the above-mentioned 3.7 million SARS-CoV-2 genomes are sorted by the sample collection date (from December 23, 2019 to May 5, 2024). For each genome

variant, both the total difference from the reference genome and the difference in the number of mutations, deletions and insertions were calculated. Based on these data, changes in the virus population were visualized over time, including the disappearance of old and the emergence of new genetic lines, emerging due to deletions, insertions and recombinations. As a result, without explicitly constructing a phylogenetic tree, an analogue of it was obtained, showing in detail the ongoing evolutionary processes, including periods of slow changes caused by mutations (occurring at almost similar rate, approximately 0.044 mutations per day, for all these years), and rapid jumps over longer distances through recombinations, deletions and insertions.

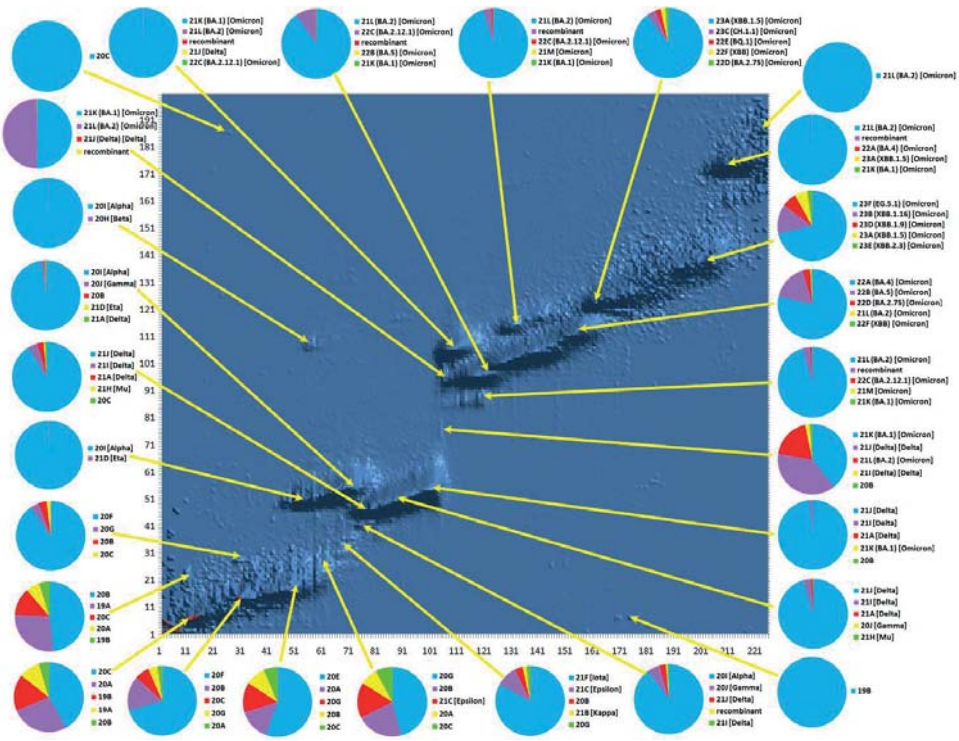


Fig. 2. Landscape of variants of SARS-CoV-2, "visited" by virus since its appearance in the end of 2019 till nowadays. Horizontal axis is time (in weeks) since 23.12.2019, vertical axis – editing distance between considered variant and reference sequence (Wuhan-Hu-1/2019)

Conclusion: A powerful, fast and efficient software tool, ParSeq v.1.1, has been developed for working with large volumes of genetic data of SARS-CoV-2 viral sequences, which has already yielded a number of interesting results and also opens up prospects for new research. In addition, the proposed approach can be applied not only to coronaviruses, but also to many other viral genomes.

Funding: The study is supported by RSF grant (No. 23-64-00005).

References

1. Aksamentov I., Roemer C. et al. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J. Open Source Software. 2021;6(67):3773. doi 10.21105/joss.03773
2. Sayers E.W., Bolton E.E. et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2022;50(D1):D20-D26. DOI 10.1093/nar/gkab1112
3. Palyanov A.Yu., Palyanova N.V. On the space of SARS-CoV-2 genetic sequence variants. *Vavilov Journal of Genetics and Breeding*. 2023;27(7): 839-850. doi 10.18699/VJGB-23-97

doi 10.18699/bgrs2024-12.3-23

AthRiboNC: unveiling the translational potential of Arabidopsis ncRNAs

Shen Y., Chen M.*

Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou, China
** mchen@zju.edu.cn*

Key words: ncRNAs; sORFs; Ribo-seq; *Arabidopsis thaliana*; coding potential

Motivation and Aim: Non-coding RNAs (ncRNAs) play significant regulatory roles in gene expression. Recent studies have discovered that some ncRNAs harbor small open reading frames (sORFs) with the potential to encode active peptides. However, research on ncRNA-encoded peptides in plants is not extensive, and there is a lack of systematic databases. Therefore, we developed AthRiboNC, which provides evidence related to Arabidopsis ncRNA-encoded peptides, with the hope of offering bioinformatics guidance for wet lab experimental research.

Methods and Algorithms: We collected Arabidopsis Ribo-seq data and sample processing metadata from 226 samples across 26 different BioProjects from the GEO and SRA databases [1]. The Ribo-seq data were aligned to the TAIR10 reference genome [2] and the PlantcircBase [3] circRNA library using HISAT2 [4]. FeatureCounts [5] was used to quantify the translation levels of miRNAs, lncRNAs, and circRNAs in different samples. ORFfinder and RiboTISH [6] were employed to identify sORFs in ncRNAs with translational potential. Co-expression network analyses were carried out using the WGCNA package in R [7] to relate the translation level of ncRNAs to sample treatment and to distinguish modules. Functional analyses were based on the hub genes of the modules. Finally, the front-end and back-end of the database website were built using Django, MySQL and jQuery.

Results: AthRiboNC provides a rich catalog of Arabidopsis ncRNAs with translational potential identified from Ribo-seq data. The database includes 4869 Arabidopsis ncRNAs detected by Ribo-seq, comprising 2743 lncRNAs, 255 miRNAs, and 1871 circRNAs. Users can conveniently browse and search for Arabidopsis ncRNAs with translational potential, their sORFs, and expression levels in samples within the database. The co-expression module offers additional downstream knowledge, such as members of co-expression modules, module-trait associations, module hub genes, and functional enrichment. Additionally, the website provides a simple BLAST function to align target sequences with all sORFs in AthRiboNC. AthRiboNC is available at https://bis.zju.edu.cn/athribonc/.

Conclusions: AthRiboNC serves as a pivotal resource for exploring the translational landscape of ncRNAs in Arabidopsis. It underlines the significance of Ribo-seq data for identifying functional peptides from ncRNAs and provides a valuable tool for wet lab researchers to investigate gene function and regulatory networks.

Funding: The study is supported by the National Natural Sciences Foundation of China (No. 32070677).

References

1. Kenneth K., Shutov O., Lapoint R. et al. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res.* 2022;50(D1):D387-D390

2. Cheng C.Y., Krishnakumar V., Chan A.P. et al. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J*. 2017;89(4):789-804
3. Xu X., Du T., Mao W. et al. PlantcircBase 7.0: Full-length transcripts and conservation of plant circRNAs. *Plant Commun*. 2022;3(4):100343
4. Kim D., Paggi J.M., Park C. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907-915
5. Liao Y., Smyth G.K., Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930
6. Zhang P., He D., Xu Y. et al. Genome-wide identification and differential analysis of translational initiation. *Nat Commun*. 2017;8(1):1749
7. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559

doi 10.18699/bgrs2024-12.3-24

Analysis of the relationships of significant contextual signals in ChIP-Seq peaks

Vishnevsky O.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *oleg@bionet.nsc.ru*

Key words: motif discovery; composite elements; transcription regulation; ChIP-Seq

Motivation and Aim: The development of ChIP-seq approaches has led to the accumulation of a huge body of information about the *in vivo* genome-wide localization of thousands of transcription factors in various cellular conditions on DNA. ChIP-seq peak sequences are characterized by a high-level complexity of contextual organization and may either contain several potential binding sites for target TFs or not contain them at all. Additionally, they may contain quite a few potential binding sites for partner TFs. A wealth of experimental data provide evidence that transcription factors regulate transcription in close cooperation with each other. Thus, taking into account pairwise and groupwise interactions between transcription factors is necessary for a deeper understanding of gene regulation and for scoring the observed ChIP-seq peaks.

Methods and Algorithms: In the present work, we analyzed the contextual organization of ChIP-seq peak sequences in experiments with 10 TFs from CistromeDB [1] belonging to six main types of DNA-binding domains [2]. Using the original *de novo* GPU-based motif discovery method Argo_CUDA [3], we revealed, in the peak sequences from each ChIP-seq experiment, both sets of significant IUPAC motifs corresponding to the target TF binding sites studied in each experiment and specific sets of motifs corresponding to the binding sites for the partner TFs. Unlike heuristic methods, Argo_CUDA evaluates the significance of all possible IUPAC motifs of a given length, which guarantees finding a global optimum. The annotation of the identified motifs was carried out using the Tomtom [4] software package based on the Hocomoco [5] and Jaspar [6] positional weight matrix databases.

Results: A significant correlation was shown between the presence of motifs corresponding to the binding sites for the target transcription factors and the partner motifs corresponding to the binding sites for the partner transcription factors, which was experimentally confirmed by the scientific literature. For all the ChIP-seq experiments considered, multiple regression models were constructed, demonstrating a significant dependence of the ChIP-seq peak sequence scores on the presence of sets of specific IUPAC motifs in these sequences. It has been shown that the most significant target motifs make a substantial contribution to the observed dependence. At the same time, the prediction quality can be improved through the use of less significant motifs as well as partner motifs.

Conclusion: The contextual features of ChIP-seq peaks that we have identified can be used to set up experiments aimed at testing potential partner interactions of TFs, the motifs of which are reliably co-represented in the sequences of ChIP-Seq peaks and also help in building potential regulatory gene networks involved in subtle developmental

processes and tissue-specific gene expression. In addition, the significant IUPAC motifs we identified can be used to develop new methods for predicting the localization of potential TFBSs in genomic sequences.

Funding: The study is supported by Russian government project (No. FWNR-2022-0020).

References

1. Zheng R., Wan C., Mei S., Qin Q., Wu Q., Sun H., Chen C.H., Brown M., Zhang X., Meyer C.A., Liu X.S. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* 2019;47:D729-D735. doi 10.1093/nar/gky1094
2. Wingender E., Schoeps T., Haubrock M., Krull M., Dönitz J. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic acids res.* 2018;46:D343-D347. doi 10.1093/nar/gkx987
3. Vishnevsky O.V., Bocharnikov A.V., Kolchanov N.A. Argo_CUDA: Exhaustive GPU based approach for motif discovery in large DNA datasets. *J Bioinform Comput Biol.* 2018;16:1740012. doi 10.1142/S0219720017400121
4. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., Noble, W. S. Quantifying similarity between motifs. *Genome Boil.* 2007;8:R24. doi 10.1186/gb-2007-8-2-r24
5. Kulakovskiy I.V., Vorontsov I.E., Yevshin I.S., Sharipov R.N., Fedorova A.D., Rumynskiy E.I., Medvedeva Y.A., Magana-Mora A., Bajic V.B., Papatsenko D.A., Kolpakov F.A., Makeev V.J. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* 2018;46:D252-D259. doi 10.1093/nar/gkx1106
6. Castro-Mondragon J.A., Riudavets-Puig R., Rauluseviciute I., Lemma R.B., Turchi L., Blanc-Mathieu R., Lucas J., Boddie P., Khan A., Manosalva Pérez N., Fornes O., Leung T.Y., Aguirre A., Hammal F., Schmelter D., Baranasic D., Ballester B., Sandelin A., Lenhard B., Vandepoele K., … Mathelier A. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2022;50:D165-D173. doi 10.1093/nar/gkab1113

doi 10.18699/bgrs2024-12.3-25

InTxSEs: a platform for interactions between bacterial secreted proteins and human proteins

Zhu Y.Y., Liu L.Y., Hu Y.M., Chen M.*

Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou, China

** mchen@zju.edu.cn*

Key words: bacterial secretion system; docking; PPIs; network; 3D structure

Motivation and Aim: Protein secretion plays a central role in modulating the interactions of bacteria with their environments. Bacteria transport numerous substrates across cellular membranes via secretion systems that are essential for virulence and survival [1]. Initial bacterial attachment to host cell surface receptors is commonly followed by induction of signaling events that induce cytoskeletal rearrangements and consequently promote intimate bacteria–host association or bacterial internalization. These events are typically followed by additional manipulations of host cell signaling events such as those involved in attenuation or evasion of the host's innate defenses [2]. In past decades, low-throughput [e. g. co-immunoprecipitation (Co-IP)] and high-throughput [e. g. mass spectrometry (MS) and yeast two-hybrid (Y2H)] techniques [3–5] allowed the determination of human–bacterial PPIs on an unprecedented scale, providing an abundance of PPI data that have been stored in a series of state-of-the-art host-pathogen protein interaction databases, such as IntAct [6]. Currently, there is a lack of specialized databases for the interaction between secreted proteins in bacterial secretion systems and human proteins. Existing databases are incomplete, lacking annotations, including the display of protein 3D structures, as well as the presentation of their interaction interfaces and functional sites. Here we introduce a comprehensive platform for the interaction between bacterial secreted proteins and human proteins. This platform contains data resources for the interaction between secretion system effectors (TxSEs) from ten bacterial secretion systems and human proteins. It provides experimentally validated Protein-Protein Interactions (PPIs), 3D complex structures of protein interactions, as well as displays of the interaction interfaces and functional sites of PPIs. Additionally, we will continuously update predicted PPIs in the future.

Methods and Algorithms:

Collection of experimentally verified PPI data

We collected experimentally verified PPIs from five public databases (i. e. HPIDB [7], PHISTO [8], IntAct [6]) and recently published literature. We obtained a total of 1,853 PPIs, In particular, we mapped protein IDs from different databases to UniProt IDs or NCBI IDs, gene names or protein names as query options in InTxSEs.

Three-dimensional complex structures of PPIs

Retrieve the three-dimensional structures of bacterial secreted proteins and human proteins separately from the PDB database [9] and the AlphaFold Protein Structure Database. For each pair of interacting proteins, InTxSEs will employ ZDOCK 3.0.2 [10], HDOCK [11], and AlphaFold3 [12] to calculate the potential complex structures with the highest docking scores. We used the obtained pdb to analyze the interaction sites using PDBePISA.

Collection of other information on proteins secreted by bacteria

We used TMHMM 2.0 software [13] to predict transmembrane helix sites for bacterial secreted proteins, and SignalP 6.0 [14] to predict signal peptide sites for bacterial secreted proteins.

Construct PPI network

We obtained the secretion protein network for each bacterial species and the corresponding protein network for human proteins from STRING [15]. A JavaScript library Cytoscape [16] is employed to display 2D protein interaction networks for each bacterial species. A JavaScript libraries echarts shows each secreted protein and human protein interaction network.

Results: As a major component, the InTxSEs online resource includes a comprehensive data module and a blast tool. Currently, InTxSEs offers 1,853 experimentally validated PPIs, 193 bacterial-secreted proteins, 102 bacterial species, and the main architecture of HVIDB contains the following key features: (1) For network information, InTxSEs provides a related human-bacterial ppi subnet for each queried PPI and a related human-bacterial PPI network for each bacterial species, which can improve the understanding of the mechanism of viral infection. (2) In terms of structural information, InTxSEs provides 3D visualization of the docking structure of PPI and the corresponding interaction sites, and the interaction sites can be queried. (3) It complements the transmembrane information of bacterial secreted proteins, signal peptide site information and subcellular localization information, which can deepen the understanding of the mechanism of secreted proteins acting on human proteins. (4) InTxSEs has a blast module to analyze homologous proteins.

Fig. 1. Main search and browse interfaces of InTxSEs

Conclusion: InTxSEs is a freely accessible resource that provides comprehensive, downloadable PPI data on bacteria-secreted proteins and human proteins for maximum application. In contrast to existing human-bacterial PPI databases, InTxSEs specifically integrates multiple PPI data resources on bacteria-secreted proteins, enabling users to explore the corresponding biological applications of corresponding interaction groups. In terms of future development, we will regularly upgrade InTxSEs by integrating more data resources, developing predictive tools to add to the platform, and designing a simpler and more user-friendly interactive analytics interface. For example, we will add drug targets that act on secreted protein structures to InTxSEs, potentially accelerating drug target discovery and drug retargeting to fight deadly diseases. All in all, InTxSEs can serve as a one-stop knowledge base to understand humans' relationship with

pathogenic bacteria and develop treatments to address the ongoing challenges of bacterial infections.

Funding: The study is supported by the National Natural Sciences Foundation of China (No. 31771477, 32070677).

References

1. Tseng T.T., Tyler B.M., Setubal J.C. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol*. 2009;9(Suppl. 1):S2. doi 10.1186/1471-2180-9-S1-S2
2. Mattoo S., Lee Y.M., Dixon J.E. Interactions of bacterial effector proteins with host proteins. *Curr Opin Immunol*. 2007;19(4):392-401. doi 10.1016/j.coi.2007.06.005
3. Walch P., Selkrig J., Knodler L.A. et al. Global mapping of Salmonella enterica-host protein-protein interactions during infection. *Cell Host Microbe*. 2021;29(8):1316-1332.e12. doi 10.1016/j.chom.2021.06.004
4. Paiano A., Margiotta A., De Luca M., Bucci C. Yeast Two-Hybrid Assay to Identify Interacting Proteins. *Curr Protoc Protein Sci*. 2019;95(1):e70. doi 10.1002/cpps.70.
5. Turriziani B., von Kriegsheim A., Pennington S.R. Protein-Protein Interaction Detection Via Mass Spectrometry-Based Proteomics. *Adv Exp Med Biol*. 2016;919:383-396. doi 10.1007/978-3-319-41448-5_18
6. Del Toro N., Shrivastava A., Ragueneau E. et al. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res*. 2022;50(D1):D648-D653. doi 10.1093/nar/gkab1006
7. Ammari M.G., Gresham C.R., McCarthy F.M., Nanduri B. HPIDB 2.0: a curated database for host-pathogen interactions. *Database (Oxford)*. 2016;2016:baw103
8. Durmuş Tekir S., Çakır T., Ardiç E. et al. PHISTO: pathogen-host interaction search tool. *Bioinformatics*. 2013;29(10):1357-1358. doi 10.1093/bioinformatics/btt137
9. Rose P.W., Prlić A., Altunkaya A. et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res*. 2017;45(D1):D271-D281. doi 10.1093/nar/gkw1000
10. Pierce B.G., Wiehe K., Hwang H., Kim B.H., Vreven T., Weng Z. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*. 2014;30(12):1771-3. doi 10.1093/bioinformatics/btu097
11. Yan Y., Zhang D., Zhou P., Li B., Huang S.Y. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res*. 2017;45(W1):W365-W373. doi 10.1093/nar/gkx407
12. Abramson J., Adler J., Dunger J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024. doi 10.1038/s41586-024-07487-w
13. Krogh A., Larsson B., von Heijne G., Sonnhammer E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305(3):567-580
14. Teufel F., Almagro Armenteros J.J., Johansen A.R. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol*. 2022;40(7):1023-1025. doi 10.1038/s41587-021-01156-3
15. Szklarczyk D., Kirsch R., Koutrouli M. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 2023;51(D1):D638-D646. doi 10.1093/nar/gkac1000
16. Franz M., Lopes C.T., Huck G. et al. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*. 2016;32:309-311

13

Симпозиум «Генетика и системная биология старения»

Symposium “Genetics and systems biology of aging”

2208

doi 10.18699/bgrs2024-13-01

Роль посттранскрипционной регуляции экспрессии гена, кодирующего протеинкиназу GSK3, в контроле тканевого гомеостаза и выживания *Drosophila melanogaster*

Андреев Ю.А.*, Пасюкова Е.Г.

Национальный исследовательский центр «Курчатовский институт», Москва, Россия
** y.andreev96@yandex.ru*

Ключевые слова: гомеостаз; выживание; жировая ткань; GSK3; *Drosophila melanogaster*

Мотивация и цель: В результате вредных воздействий, а также в ходе старения происходит нарушение гомеостаза, в том числе изменяется экспрессия генов и вследствие этого внутриклеточная регуляция биохимических процессов. Контролируемое точечное нарушение гомеостаза может позволить подробнее изучить процессы, происходящие в процессе старения и развития патологий. Киназа гликогенсинтазы GSK3 (Glycogen Synthase Kinase 3) представляет собой высококонсервативный фермент, являющийся ключевым звеном множества метаболических путей. Описана важная роль GSK3 в развитии различных патологических состояний, таких как болезнь Альцгеймера, диабет, воспалительные заболевания и некоторые виды рака. Целью данной работы была оценка влияния сверхэкспрессии нормальной и измененных форм GSK3 на выживание *D. melanogaster* и сохранение гомеостаза различных тканей, таких как жировое тело и слюнные железы.

Методы и алгоритмы: У *D. melanogaster* протеинкиназу GSK3 кодирует ген *shaggy* (*sgg*). Сверхэкспрессия GSK3 достигалась путем тканеспецифической индукции трансгенов, кодирующих нормальную изоформу GSK3B (sggB) и изоформу GSK3B с аминокислотными заменами, приводящими к нарушению внутриклеточной регуляции активности GSK3 (sggB_S9A), снижению активности GSK3 (sggB_Y214F) и появлению неактивной GSK3 (sggB_A81T) [1]. Для индукции трансгенов использовали бинарную систему GAL4-UAS [2]. Сверхэкспрессию трансгенов индуцировали в жировом теле и слюнных железах личинок. Количество транскрипта *sgg* оценивали с помощью количественной ПЦР в реальном времени. Количество белка GSK3 определяли с помощью вестерн-блот анализа. Объем и количество секреторных клеток слюнных желез, площадь поперечного сечения жировых клеток определяли с помощью конфокальной микроскопии. Количество общего белка в жировом теле оценивали методом Бредфорда. Уровень гликогена в жировом теле определяли с помощью ШИК-реакции.

Результаты: В результате индукции каждого из использованных трансгенов в жировом теле личинок наблюдалось увеличение количества транскрипта *sgg*. Однако увеличение количества белка GSK3 наблюдалось только в случае сверхэкспрессии sggB_S9A. Снижение уровня гликогена в жировом теле, которое отражает увеличение активности GSK3, было выявлено также только в результате сверхэкспрессии sggB_S9A, а в случае сверхэкспрессии sggB_Y214F наблюдалось увеличение уровня гликогена. Количество общего белка в жировом теле, объем

слюнных желез при отсутствии изменения количества секреторных клеток снижались во всех случаях. Площадь поперечного сечения жировых клеток снижалась только в случае дополнительной экспрессии sggB и sggB_S9A. Более того, в результате сверхэкспрессии как sggB, так и sggB_S9A в жировом теле и слюнных железах максимальная продолжительность жизни была ограничена куколочной стадией p5(ii). При сверхэкспрессии sggB_Y214F и sggB_A81T развитие протекало нормально, а продолжительность жизни взрослых особей, как было показано ранее [3], изменялась.

Обсуждение и выводы: Отсутствие увеличения количества белка GSK3 при возросшем в результате индукции трансгенов количестве транскрипта гена *sgg* в жировом теле может быть связано с посттранскрипционной регуляцией экспрессии трансгенов клеточными факторами. Обращает на себя внимание тот факт, что сверхэкспрессия sggB_S9A, если и регулируется посттранскрипционно, то в меньшей степени, чем остальные трансгены. Показано, что в клетках эпителиальной опухоли молочной железы замена серина на аланин в 9 положении (S9A) GSK3 приводит к нарушению протеасомной деградации этого белка [4]. Можно предположить, что и в наших опытах по крайней мере одним из механизмов постранскрипционной регуляции количества GSK3 является протеасомная деградация, которая не затрагивает белок с заменой S9A. Известно, что замена S9A нарушает регуляцию активности GSK3 другими протеинкиназами клетки. Фосфорилирование серина 9 делает возможным взаимодействие N-концевого участка GSK3 с сайтом ее связывания с праймированными сериновыми остатками белков-мишеней. В результате сайт связывания и, следовательно, активность GSK3 в отношении мишеней, требующих праймирования, могут быть блокированы. Замена серина на аланин приводит к тому, что GSK3 сохраняет постоянную активность в отношении таких мишеней, к которым относится и гликогенсинтаза. Тот факт, что снижение уровня гликогена в жировом теле было выявлено только при сверхэкспрессии sggB_S9A, но не других форм, можно объяснить как увеличением количества GSK3, которое наблюдалось только в этом случае, так и конститутивной активностью этой формы белка. Увеличение количества гликогена в результате сверхэкспрессии sggB_Y214F указывает на снижение суммарной внутриклеточной активности GSK3 в отношении гликогенсинтазы в этом случае. Известно, что фосфорилирование тирозина в положении 214 увеличивает активность GSK3 [5], следовательно, замена Y214F снижает активность белка, кодируемого трансгеном. Снижение суммарной активности GSK3 в клетке подтверждает ранее высказанное предположение о том, что трансген sggB_Y214F может проявлять отдельные свойства конструкции с доминантным негативным эффектом [1]. Однако причина того, почему сверхэкспрессия sggB_A81T, эффект которой тоже, предположительно, является доминантно негативным [1], не оказала подобного влияния на уровень гликогена, остается не ясной.

Известно, что протеинкиназа GSK3 фосфорилирует и тем самым инактивирует фактор инициации трансляции eIF6 [6], что может объяснять снижение количества общего белка в жировом теле. Фосфорилирование eIF6 не требует праймирования [6], поэтому эффект активации экспрессии всех трансгенов оказался одинаковым. Уменьшение размера клеток жирового тела, а также объема слюнных желез можно объяснить влиянием сверхэкспрессии *sgg* на активность TOR пути [7], участвующего в контроле роста клеток. Причины различного влияния

сверхэкспрессии разных трансгенов на выживание на куколочной стадии неясны и требуют дальнейшего изучения. Однако прежде всего необходимо понять, почему совокупность эффектов сверхэкспрессии *sgg* наблюдается при отсутствии детектируемого увеличения количества белка GSK3. Возможно, для этого достаточно очень небольшого избытка GSK3, который не удается выявить методом вестерн-блот анализа.
Финансирование: Работа выполнена при финансовой поддержке государственного бюджета Национального исследовательского центра «Курчатовский институт».

Impact of post-transcriptional regulation of expression of the gene encoding GSK3 protein kinase in the control of tissue homeostasis and survival in *Drosophila melanogaster*

Andreev Y.A.*, Pasyukova E.G.
National Research Centre "Kurchatov Institute", Moscow, Russia
** y.andreev96@yandex.ru*

Key words: homeostasis; survival; adipose tissue; GSK3; *Drosophila melanogaster*

Motivation and Aim: Exposure to harmful stimuli as well as aging can disrupt homeostasis, leading to changes in gene expression and subsequent intracellular regulation of biochemical processes. Controlled disruption of homeostasis could provide a more detailed understanding of the processes occurring during aging and the development of pathologies. Glycogen synthase kinase (GSK3) is a highly conserved enzyme that plays a key role in numerous metabolic pathways. GSK3 has been implicated in the development of various pathological conditions such as Alzheimer's disease, diabetes, inflammatory diseases, and certain types of cancer. The aim of this study was to assess the impact of overexpression of normal and altered forms of GSK3 on the survival of *D. melanogaster* and the maintenance of homeostasis in different tissues such as the fat body and salivary glands.
Methods and Algorithms: In *D. melanogaster*, the protein kinase GSK3 is encoded by the gene *shaggy* (*sgg*). Overexpression of GSK3 was achieved through tissue-specific induction of transgenes encoding the normal isoform of GSK3B (sggB) and isoform GSK3B with amino acid substitutions leading to disruption of intracellular regulation of GSK3 activity (sggB_S9A), reduction in GSK3 activity (sggB_Y214F) and generation of inactive GSK3 (sggB_A81T) [1]. Transgenes were induced using the binary GAL4-UAS system [2]. Overexpression of transgenes was induced in the fat body and salivary glands of larvae. The amount of *sgg* transcript was evaluated using quantitative real-time PCR. The level of GSK3 protein was determined using Western blot analysis. The volume and number of secretory cells in the salivary glands, cross-sectional area of fat cells were determined using confocal microscopy. Total protein content in the fat body was assessed using the Bradford method. The level of glycogen in the fat body was determined using the PAS (Periodic Acid-Schiff) reaction.
Results: Induction of each of the transgenes used in the fat body of larvae resulted in an increase in the amount of *sgg* transcript. However, an increase in the level of GSK3 protein was observed only in the case of overexpression of sggB_S9A. A decrease in the level of glycogen in the fat body, reflecting an increase in GSK3 activity, was also detected only as a result of overexpression of sggB_S9A, while overexpression of sggB_Y214F led to

an increase in the level of glycogen. The total protein content in the fat body, and the volume of the salivary glands, in the absence of changes in the number of secretory cells, decreased in all cases. The cross-sectional area of fat cells was reduced only in the case of additional expression of sggB and sggB_S9A. Furthermore, as a result of overexpression of both sggB and sggB_S9A in the fat body and salivary glands, the maximum lifespan was limited to the p5(ii) pupal stage. With overexpression of sggB_Y214F and sggB_A81T, development proceeded normally, but the lifespan of adult individuals, as previously shown [3], was altered.

Discussion and Conclusions: The lack of an increase in the level of GSK3 protein with the increased amount of *sgg* gene transcript induction in the fat body may be related to post-transcriptional regulation of transgene expression by cellular factors. It is noteworthy that the overexpression of sggB_S9A, if post-transcriptionally regulated, appears to be to a lesser extent than the other transgenes. It has been demonstrated that in breast epithelial tumor cells, the substitution of serine with alanine at position 9 (S9A) in GSK3 disrupts proteasomal degradation of this protein [4]. It can be hypothesized that, in our experiments, at least one of the mechanisms of post-transcriptional regulation of GSK3 levels could involve proteasomal degradation, which does not affect the protein with the S9A substitution. It is known that the S9A substitution disrupts the regulation of GSK3 activity by other protein kinases within the cell. Phosphorylation of serine 9 enables the N-terminal region of GSK3 to interact with the site of primed serine residues of target proteins. As a result, the binding site and hence the activity of GSK3 towards primed targets may be blocked. The substitution of serine with alanine leads to GSK3 maintaining a constant activity towards such targets, including glycogen synthase. The fact that a decrease in the level of glycogen in the fat body was observed only with the overexpression of sggB_S9A, but not with other forms, can be explained by both the increase in the amount of GSK3, which was only observed in this case, and the constitutive activity of this form of the protein. Increase in glycogen levels resulting from overexpression of sggB_Y214F indicates a decrease in the overall intracellular activity of GSK3 towards glycogen synthase in this case. It is known that phosphorylation of tyrosine at position 214 enhances GSK3 activity [5], therefore, the substitution Y214F reduces the activity of the protein encoded by the transgene. The decrease in the total activity of GSK3 in the cell confirms the previously stated assumption that the sggB_Y214F transgene may exhibit individual properties of the construct with a dominant negative effect [1]. However, the reason why overexpression of sggB_A81T, the effect of which is also presumably dominant negative [1], did not have a similar effect on glycogen levels remains unclear.

It is known that the protein kinase GSK3 phosphorylates and thereby inactivates the translation initiation factor eIF6 [6], which may explain the decrease in the total protein content in the fat body. Phosphorylation of eIF6 does not require priming [6], hence the effect of activation of the expression of all transgenes was identical. The reduction in the size of fat body cells and salivary gland volume can be explained by the influence of *sgg* overexpression on the activity of the TOR pathway [7], which is involved in cell growth control. The reasons for the different effects of overexpression of different transgenes on survival at the pupal stage are unclear and require further investigation. However, it is essential to understand first and foremost why the combination of effects of *sgg* overexpression is observed in the absence of detectable increase in GSK3 protein levels. It is possible that this could be due to a very slight excess of GSK3, which cannot be detected by Western blot analysis.

Funding: The work was carried out within the state assignment of NRC "Kurchatov Institute".

doi 10.18699/bgrs2024-13-02

Уровень экспрессии GSK3 в отдельных группах дофаминергических нейронов влияет на продолжительность жизни и поведение *Drosophila melanogaster*

Веселкина Е.[1]*, Рощина Н.[1, 2], Тростников М.[3], Пасюкова Е.[1]

[1] *НИЦ «Курчатовский институт», Москва, Россия*

[2] *Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия*

[3] *Институт биологии гена РАН, Москва, Россия*

* *veselkinaer123@gmail.com*

Ключевые слова: продолжительность жизни; плодовитость; пищевое поведение; дофаминергические нейроны; *Drosophila melanogaster*

Мотивация и цель: Дофаминергические (ДА) нейроны являются важной частью нервной системы. И у *Drosophila melanogaster,* и у млекопитающих ДА нейроны выполняют сходные функции: участвуют в контроле: пищевого поведения, подвижности, циркадных ритмов, памяти, формировании мотивации. В течение жизни дофаминергическая система изменяется. Даже у здоровых людей к пожилому возрасту наблюдается снижение количества дофамина и ДА нейронов. С потерей ДА нейронов связаны многие нейродегенеративные заболевания [1,2]. Дофаминергическая система *D. melanogaster* устроена относительно просто и включает в себя всего 125 ДА нейронов на полушарие головного мозга; клеточные тела ДА нейронов сгруппированы в кластеры и легко визуализируются [3,4]. Относительная простота устройства дофаминергической системы *D. melanogaster*, при функциональной схожести с дофаминергической системой млекопитающих, делает ее прекрасным модельным объектом для изучения вклада отдельных генов в контроль различных свойств и функций ДА нейронов. Мы исследовали роль ДА нейронов в контроле продолжительности жизни (ПЖ).

Важное значение в поддержании стабильности дофаминергической системы играет ген *shaggy (sgg)*, кодирующий высоконсервативную протеинкиназу киназу гликогенсинтазы 3 (Glycogen Syntase Kinase 3, GSK3). GSK3 фосфорилирует транскрипционные факторы CREB и NURR1 [5,6], которые, в свою очередь, регулируют работу гена *pale*, кодирующего тирозингидроксилазу, ключевой фермент биосинтеза дофамина.

Ранее мы показали, что эктопическая экспрессия дополнительной копии гена *shaggy*, кодирующей основной транскрипт sggRB, во всех ДА нейронах приводила к летальному эффекту. Парадоксальным образом, экспрессия той же дополнительной копии гена *shaggy* во всей нервной системе, включая ДА нейроны, не вызывала летальности, что может говорить о существовании непонятных пока компенсаторных эффектов, вызванных увеличением экспрессии *shaggy* в других типах нейронов. Эти данные кажутся еще более интересными ввиду того, что эктопическая экспрессии дополнительной копии основного транскрипта с однонуклеотидной заменой в каталитическом центре, sggRB-A81T, приводящая к снижению активности фермента в клетке вследствие доминантно

негативного эффекта [7,8], вызывала статистически значимое увеличение ПЖ самок.

В последнее время все больше данных свидетельствует о том, что отдельные нервные клетки могут определять те или иные специфические типы поведения. Это свидетельствует о существенной дифференцировке и специализации нервных клеток, в том числе, нейронов одной эргичности. Перед нами встала задача понять, может ли изменение в работе одного гена, гена *shaggy*, в отдельных группах ДА нейронов привести к изменениям ПЖ, повлиять на функциональное состояние нервной системы и поведение.

Материалы и методы: Для оценки влияния экспрессии sggRB-A81T в отдельных кластерах ДА нейронов на функциональное состояние нервной системы и ПЖ мы использовали модификацию классической бинарной экспрессионной системы GAL4-UAS [9], Split_GAL4, которая позволяет индуцировать экспрессию конструкции только в отдельных группах дофаминергических неронов [10]. ПЖ и количество съеденной пиши оценивали, как описано в [11]. Визуализацию нейронов проводили с помощью конфокального микроскопа, GAL4-UAS-GFP и антител к nc82. Интерес к пище оценивали, как описано в [12].Количество съеденной пищи учитывали, как описано[13]. Статистическую обработку всех данных проводили стандартными методами [11].

Результаты и обсуждение: Мы оценили влияние измененной экспресии *shaggy*, которая обеспечивала доминантно негативный эффект в отдельных кластерах ДА нейронов на ПЖ девственных самок и самцов *D. melanogaster*. У самцов влияние на ПЖ оказывает экспрессия в кластере pal, достоверно снижая ее. У самок изменение экспрессии *shaggy* в кластере ppm2 приводит к увеличению ПЖ девственных самок, но этот эффект проявляется только в отсутствие экспрессии *shaggy* в других кластерах, что может говорить о разнонаправленном вкладе разных групп ДА нейронов в контроль ПЖ. ДА нейроны гетерогенны. Это связано, с одной стороны, с тем, куда проецируется определенный нейрон, а с другой стороны, с тем, что нейроны значительно отличаются по уровню синтеза дофамина и, следовательно, по функциональному статусу. Поскольку синтез дофамина зависит от активности GSK3, можно предположить, что активность этой протеинкиназы различна в разных ДА нейронах. Тогда экспрессия одной и той же дополнительной копии в разных ДА нейронах может оказывать разное воздействие на их функцию и, далее, на ПЖ. В то же время, возможно, что характер влияния на ПЖ зависит от того, куда проецируются нейроны.

Известно, что дофамин может выполнять две функции, он является нейромедиатором и гормоном. ДА нейроны сильно связаны с репродуктивной функцией [14]. Мы выяснили, что у скрещенных самок изменение экспрессии *shaggy* в отдельных кластерах ДА нейронов приводит к увеличению ПЖ. Чтобы ответить на вопрос, не связано ли это с эффектом trade off, когда увеличение продолжительности жизни приводит к снижению плодовитости, мы оценили плодовитость самок на 5, 30 и 50 день их жизни. Оказалось, что измененная экспрессия shaggy в различных кластерах ДА нейронов не влияет на репродуктивную функцию, а в некоторых случаях даже улучшает ее. Таким образом, локальные изменения экспрессии *shaggy* в кластерах ДА нейронов приводят к увеличению ПЖ самок, не снижая их плодовитости.

ДА нейроны определяют характер пищевого поведения и особенности обоняния мух. Мы предположили, что увеличение ПЖ связано с ограничением в

потреблении калорий из-за потери интереса к пище у мух с измененной экспрессией *shaggy* в отдельных кластерах ДА нейронов. Однако мы не обнаружили снижения интереса к пище у мух с повышенной ПЖ, а дополнительная экспрессия sggRB-A81T в нейронах кластера pal приводила даже к увеличению потребления пищи.
Выводы: Мы доказали, что изменения в работе одного гена в отдельных кластерах ДА нейронов могут влиять на ПЖ. Мы также еще раз продемонстрировали половой диморфизм в контроле ПЖ: ПЖ самцов и самок зависит от разных ДА нейронов с измененной экспрессией *shaggy*. Изменение ПЖ не объясняется ни компенсаторным ослаблением репродуктивной функции, ни снижением потребления пищи и, вследствие этого, уменьшения калорийности питания. Поиск молекулярных и клеточных механизмов влияния GSK3 на ПЖ будет продолжен.
Финансирование: Работа проведена в рамках выполнения государственного задания НИЦ «Курчатовский институт».

The expression level of GSK3 in specific groups of dopaminergic neurons influences the lifespan and behavior of *Drosophila melanogaster*

Veselkina E.[1]*, Roshina N.[1, 2], Trostnikov M.[3], Pasyukova E.[1]

[1] *National Research Center "Kurchatov Institute", Moscow, Russia*

[2] *N. I. Vavilov Institute of General Genetics, RAS, Moscow, Russia*

[3] *Institute of Gene Biology, RAS, Moscow, Russia*

* *veselkinaer123@gmail.com*

Key words: lifespan; fertility; feeding behaviour; dopaminergic neurons; *Drosophila melanogaster*

Motivation and Aim: Dopaminergic (DA) neurons are an important part of the nervous system. In both *Drosophila melanogaster* and mammals, DA neurons perform similar functions: they are involved in regulating feeding behavior, mobility, circadian rhythms, memory, and motivation. The dopaminergic system changes throughout life. Even in healthy individuals, a decrease in the amount of dopamine and DA neurons is observed in old age. Loss of DA neurons is associated with many neurodegenerative diseases. The dopaminergic system of *D. melanogaster* is relatively simple, consisting of only 125 DA neurons per hemisphere of the brain; the cell bodies of DA neurons are clustered and easily visualized. The relative simplicity of the dopaminergic system in *D. melanogaster*, along with its functional similarity to the dopaminergic system of mammals, makes it an excellent model for studying the contribution of individual genes in controlling various properties and functions of DA neurons. We investigated the role of DA neurons in regulating lifespan.
The gene *shaggy* (*sgg*), encoding the highly conserved protein kinase glycogen synthase kinase 3 (GSK3), plays a crucial role in maintaining the stability of the dopaminergic system. GSK3 phosphorylates the transcription factors CREB and NURR1, which in turn regulate the function of the pale gene, encoding tyrosine hydroxylase, a key enzyme in dopamine biosynthesis.
Previously, we have shown that the ectopic expression of an additional copy of the shaggy gene, encoding the main transcript sggRB, in all DA neurons led to a lethal effect.

Paradoxically, the expression of the same additional copy of the shaggy gene in the entire nervous system, including DA neurons, did not cause lethality, indicating the presence of as yet unknown compensatory effects resulting from increased shaggy expression in other types of neurons. These findings are even more intriguing considering that ectopic expression of an additional copy of the main transcript with a single-nucleotide replacement in the catalytic center, sggRB-A81T, resulting in reduced enzyme activity within the cell due to a dominant negative effect, led to a statistically significant increase in the lifespan of females.

Recently, an increasing amount of data suggests that individual nerve cells can determine specific types of behavior. This indicates significant differentiation and specialization of nerve cells, including neurons of the same type. The task before us is to understand whether changes in the function of a single gene, the shaggy gene, in specific groups of DA neurons can lead to changes in DA, affect the functional state of the nervous system, and behavior.

Methods and Algorithms: To assess the impact of sggRB-A81T expression in specific clusters of DA neurons on the functional state of the nervous system and behavior, we used a modification of the classical binary expression system GAL4-UAS [9], called Split_GAL4, which allows for the induction of construct expression only in specific groups of dopaminergic neurons [10]. Lifespan and the amount of consumed food were evaluated as described in [11]. Neuron visualization was performed using a confocal microscope, GAL4-UAS-GFP, and antibodies to nc82. Interest in food was assessed as described in [12]. Food consumption was also assessed [13]. Statistical analysis of all data was conducted using standard methods [11].

Results and discussion: We evaluated the impact of altered *shaggy* expression, which provided a dominantly negative effect in specific clusters of DA neurons, on the lifespan of virgin female and male *D. melanogaster*. In males, the influence on lifespan is seen through expression in the pal cluster, significantly reducing it. In females, the alteration of shaggy expression in the ppm2 cluster leads to an increase in the lifespan of virgin females, but this effect is only evident in the absence of shaggy expression in other clusters, suggesting a diverse contribution of different neuron groups to lifespan control. DA neurons are heterogeneous. This is due, on one hand, to where a specific neuron projects, and on the other hand, to the significant differences in dopamine synthesis levels and, therefore, functional status among neurons. Since dopamine synthesis depends on GSK3 activity, it can be assumed that the activity of this protein kinase varies in different DA neurons. Therefore, the expression of the same additional copy in different DA neurons may have different effects on their function and, subsequently, on lifespan. At the same time, it is possible that the nature of the influence on DA content depends on where the neurons project.

It is known that dopamine can perform two functions, it acts as a neurotransmitter and a hormone. DA neurons are strongly linked to reproductive function [14]. We found that in crossed females, changes in *shaggy* expression in individual clusters of DA neurons lead to increased lifespan. To answer the question of whether this is not related to a trade-off effect, where increased lifespan leads to decreased fertility, we evaluated the fertility of females on days 5, 30, and 50 of their lives. It turned out that the altered expression of *shaggy* in different clusters of DA neurons does not affect reproductive function and in some cases even improves it. Thus, local changes in *shaggy* expression in DA neuron clusters lead to increased lifespan in females without reducing their fertility.

DA neurons determine the character of the feeding behavior and olfactory sensitivity of flies. We hypothesized that the increase in lifespan is associated with a limitation in calorie consumption due to a loss of interest in food in flies with altered *shaggy* expression in individual clusters of DA neurons. However, we did not find a decrease in interest in food in flies with increased lifespan, and additional expression of sggRB-A81T in the pal cluster neurons even led to an increase in food consumption.

Conclusion: We have demonstrated that changes in the function of a single gene in individual clusters of DA neurons can affect lifespan. We have also once again shown sexual dimorphism in lifespan control: the lifespan of males and females depends on different DA neurons with altered *shaggy* expression. The change in lifespan is not explained by compensatory weakening of reproductive function, nor by reduced food consumption and, as a result, reduced caloric intake. The search for molecular and cellular mechanisms of the influence of GSK3 on lifespan will continue.

Funding: The work was carried out as part of the fulfillment of the state assignment of the NRC Kurchatov Institute.

Список литературы/References

1. Morgan D.G., May P.C., Finch C.E. Dopamine and serotonin systems in human and rodent brain: effects of age and neurodegenerative disease. *J Am Geriatr Soc*. 1987;35(4):334-345
2. Zhou Z.D. et al. Role of dopamine in the pathophysiology of Parkinson's disease. *Transl Neurodegener*. 2023;12(1):44
3. Mao Z., Davis R.L. Eight different types of dopaminergic neurons innervate the Drosophila mushroom body neuropil: anatomical and physiological heterogeneity. *Front Neural Circuits*. 2009;3:5
4. Nässel D.R., Elekes K. Aminergic neurons in the brain of blowflies and Drosophila: dopamine- and tyrosine hydroxylase-immunoreactive neurons and their relationship with putative histaminergic neurons. *Cell Tissue Res*. 1992;267(1):147-167
5. Ilouz R. et al. Identification of novel glycogen synthase kinase-3beta substrate-interacting residues suggests a common mechanism for substrate recognition. *J Biol Chem*. 2006;281(41):30621-30630
6. García-Yagüe Á.J. et al. α-Synuclein Induces the GSK-3-Mediated Phosphorylation and Degradation of NURR1 and Loss of Dopaminergic Hallmarks. *Mol Neurobiol*. 2021;58(12):6697-6711
7. Bourouis M. Targeted increase in shaggy activity levels blocks wingless signaling. *Genesis*. 2002;34(1-2):99-102
8. Trostnikov M.V. et al. Modulated Expression of the Protein Kinase GSK3 in Motor and Dopaminergic Neurons Increases Female Lifespan in Drosophila melanogaster. *Front Genet Frontiers.* 2020;11:668
9. Brand A.H., Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*. 1993;118(2):401-415
10. Xie T. et al. A Genetic Toolkit for Dissecting Dopamine Circuit Function in Drosophila. *Cell Rep*. 2018;23(2):652-665
11. Tsybul'ko E. et al. The Mitochondria-Targeted Plastoquinone–Derivative SkQ1 Promotes Health and Increases Drosophila melanogaster Longevity in Various Environments. *J Gerontol A Biol Sci Med Sci*. 2017;72(4):499-508
12. Landayan D., Feldman D.S., Wolf F.W. Satiation state-dependent dopaminergic control of foraging in Drosophila: 1. *Sci Rep*. 2018;8(1):5777
13. Roshina N.V. et al. Embryonic expression of shuttle craft, a Drosophila gene involved in neuron development, is associated with adult lifespan. *Aging (Albany NY)*. 2014;6(12):1076-1093
14. Kuo S.-Y. et al. PPL2ab neurons restore sexual responses in aged Drosophila males through dopamine: 1. *Nat Commun*. 2015;6(1):7490

doi 10.18699/bgrs2024-13-03

Старение растений и глутаматдегидрогеназа: нарушение индуцированного темнотой старения у растений *Arabidopsis thaliana* L. (Heyn) мутантной линии *gdh1gdh2*

Гарник Е.Ю.[1]*, Власова А.А.[2], Вильянен Д.В.[3], Бельков В.И.[1], Тарасенко В.И.[1], Константинов Ю.М.[1, 2]

[1] *Сибирский институт физиологии и биохимии растений СО РАН, Иркутск, Россия*

[2] *Иркутский государственный университет, Иркутск, Россия*

[3] *Институт фундаментальных проблем биологии РАН, Пущинский научный центр биологических исследований РАН, Пущино, Россия*

* *elga74@yandex.ru*

Ключевые слова: *Arabidopsis thaliana*; старение растений; глутаматдегидрогеназа; индукция старения

Мотивация и цель: Модульный принцип строения тела высших растений [1, 2] обусловливает необходимость различать старение организма растения в целом и старение отдельных органов, поскольку каждый модуль/орган тела растения имеет собственный онтогенез и собственные сроки наступления разных стадий развития, включая старение. Старение листьев растений, по сути, сводится к организованному, генетически запрограммированному и строго регулируемому перераспределению ресурсов (в первую очередь – биологически доступного азота) от стареющего листа к репродуктивным, запасающим либо просто наиболее молодым органам [3]. Старение растений может быть индуцировано искусственно, причем факторы, индуцирующие начало массированного старения листьев или всего организма, часто видоспецифичны. Всё это делает растения интересной моделью для изучения генетически запрограммированных механизмов старения.

Нашей группой было обнаружено, что у растений двойного нокаут-мутанта *Arabidopsis thaliana* L. (Heyn) по генам глутаматдегидрогеназы, линия *gdh1gdh2,* при экспозиции в темноте более четырех суток (условия, индуцирующие старение у растений данного вида) не начинается пожелтение листьев, которое в норме является первым видимым признаком их старения. Глутаматдегидрогеназа (ГДГ, КФ 1.4.1.2) – митохондриальный фермент, катализирующий превращение L-глутамат в 2-оксоглутарат, а также (в некоторых условиях) обратную реакцию. Таким образом, ГДГ обеспечивает связь между метаболизмом углерода и азота благодаря катаболизму 2-оксоглутарата в цикле трикарбоновых кислот. Экспрессия генов ГДГ у растений зависит от освещенности [4]. На свету экспрессия генов и ферментативная активность ГДГ минимальны у растений дикого типа и отсутствуют у растений мутантной линии *gdh1gdh2*. В условиях длительной экспозиции в темноте экспрессия генов и активность фермента ГДГ у растений дикого типа максимальны, у мутантных отсутствуют.

Целью настоящей работы было исследование нарушений нормального развития индуцированного темнотой старения у мутанта арабидопсиса по генам ГДГ и выяснение роли глутаматдегидрогеназы в регуляции развития старения растений.

Методы и алгоритмы: После экспозиции мутантных и контрольных растений в темноте оценивали содержание и состав хлорофиллов (спектрофотометрия), содержание растворимых белков (спектрофотометрия), ионную проницаемость клеточных мембран (кондуктометрия), а также экспрессию различных групп генов, ассоциированных со старением (транскрипционные факторы, инициирующие старение; гены, контролирующие деградацию хлорофиллов; гены, обеспечивающие ремобилизацию азота и перераспределение продуктов катаболизма). Исследование экспрессии проводили методом ПЦР после обратной транскрипции. Оценку указанных выше параметров проводили также на фоне обработки растений абсцизовой кислотой (главный гормональный регулятор старения у растений).
Результаты: Исходное содержание хлорофилла не различалось у растений дикого типа и мутантной линии *gdh1gdh2*, однако деградация хлорофилла в течение семи дней в темноте у мутантных растений происходит значительно медленнее. Соотношение хлорофиллов *a/b* в розетках растений *gdh1gdh2* снижалось в течение семи дней в темноте из-за более медленной деградации хлорофилла *b*. В розетках Col-0 не наблюдалось существенного изменения соотношения хлорофиллов *a/b*.
Кондуктометрический тест на ионную проницаемость мембран отражает целостность клеточной мембраны и жизнеспособность растения в целом. Жизнеспособность растений снижалась в ходе темновой экспозиции для обоих генотипов, но гораздо быстрее для растений линии *gdh1gdh2*. Это признак фенотипа «stay-green» типа D – накопление токсичных метаболитов.
Для изучения динамики изменения экспрессии генов при нормальном течении темнового старения у растений дикого типа отбирали образцы листьев по окраске: желтые (самые нижние), желто-зеленые (средние) и зеленые (верхушечные). Пять генов деградации хлорофилла показали значительное увеличение уровней транскриптов от самых молодых (зеленых) до самых старых (полностью желтых) листьев. Для сравнения растений дикого типа и растений *gdh1gdh2* брали только нижнюю (самую старую) пару листьев из каждой розетки. Листья дикого типа демонстрировали постепенное повышение уровней транскриптов генов деградации хлорофилла на протяжении всего эксперимента. У мутантных растений экспрессия всех изученных генов деградации хлорофилла также имела тенденцию к повышению после двух дней пребывания в темноте, но дальнейшего увеличения экспрессии не наблюдалось.
Сбои в гормональной регуляции часто являются причиной нарушений старения растений, поэтому мы исследовали чувствительность растений к абсцизовой кислоте (АБК) как к основному и очень сильному индуктору старения. Чувствительность АБК у семян *gdh1gdh2* не была нарушена. Однако у взрослых растений мы наблюдали индукцию старения у растений дикого типа и гибель без признаков старения у растений *gdh1gdh2*. Мы исследовали экспрессию *SAG*s (генов, ассоциированных со старением) в верхних и нижних листьях растений после четырех дней экспозиции в темноте. В большинстве случаев индукции *SAG*s у мутантных растений не наблюдали. После обработки АБК наблюдали некоторый ответ генов *SAG113, SAG29, GLN1.1* у мутантных растений, но в целом экспрессия *SAG*s в растениях линии *gdh1gdh2* отличалась от паттернов экспрессии у растений дикого типа. Таким образом, развитие старения *gdh1gdh2* не полностью нечувствительно к АБК, однако его гормональная регуляция существенно нарушена.

Выводы: Обнаруженные факты позволяют предположить существование функциональной связи между экспрессией генов НАД(Н)-зависимой глутаматдегидрогеназы и регуляцией процессов старения у растений. Отсутствие активности глутаматдегидрогеназы при длительной темновой экспозиции, вероятно, является причиной нарушения старения растений *gdh1gdh2,* так как может приводить к переключениям метаболических путей с последующим появлением токсичных метаболитов. Таким образом, глутаматдегидрогеназа может быть одним из важных регуляторов старения арабидопсиса.
Финансирование: Исследование поддержано Министерством науки и образхзования Российской Федерации, номер государственного задания FWSS-2022-0005.

Plant senescence and glutamate dehydrogenase: a disruption of dark-induced senescence in *Arabidopsis thaliana* L. (Heyn) mutant line *gdh1gdh2*

Garnik E.Yu.[1]*, Vlasova A.A.[2], Vilyanen D.V.[3], Belkov V.A.[1], Tarasenko V.I.[1], Konstantinov Yu.M.[1, 2]

[1] *Siberian Institute of Plant Physiology and Biochemistry, SB RAS, Irkutsk, Russia*

[2] *Irkutsk State University, Irkutsk, Russia*

[3] *Institute of Basic Biological Problems, RAS, Federal Research Center, Pushchino Scientific Center for Biological Research RAS, Pushchino, Russia*

** elga74@yandex.ru*

Key words: *Arabidopsis thaliana*; plant senescence; glutamate dehydrogenase; induced senescence

Motivation and Aim: The modular body structure of higher plants [1, 2] makes it necessary to distinguish between the aging of the plant organism as a whole and the senescence of individual organs, since each module/organ of the plant body has its own ontogenesis and its own timing of the onset of different stages of development, including senescence. Senescence of plant leaves, in essence, comes down to an organized, genetically programmed and strictly regulated redistribution of resources (primarily biologically available nitrogen) from the senescing leaf to reproductive, storage, or simply the youngest organs [3]. Plant senescence can be induced artificially, and the factors inducing the onset of massive senescence of leaves or the entire organism are often species-specific. All this makes plants an interesting model for studying genetically programmed mechanisms of senescence.

Our group found that in the double knockout mutant plants *gdh1gdh2* of *Arabidopsis thaliana* L. (Heyn) carrying mutations in two glutamate dehydrogenase genes, when exposed to darkness for more than four days (conditions that induce senescence in Arabidopsiss), leaves do not turn yellow, which is normally the first visible sign of their senescence. Glutamate dehydrogenase (GDH, EC 1.4.1.2) is a mitochondrial enzyme interconverting L-glutamate and 2-oxoglutarate (forward and back reactions), and thus providing a link between carbon and nitrogen metabolism because of catabolizing of 2-oxoglutarate in the TCA cycle. *GDH* genes expression is light dependent [4]. In the light GDH expression and enzyme activity is minimal in wild-type plants and absent in

gdh1gdh2. But in the prolonged dark GDH expression and enzyme activity in wild-type plants is maximal, and still missing in the *gdh1gdh2* line plants.
The aim of this work was to study disturbances in the normal development of dark-induced senescence in the *GDH* gene mutant of Arabidopsis and to elucidate the role of glutamate dehydrogenase in the regulation of the development of plant senescence.
Methods and Algorithms: After exposure of mutant and control plants to the dark, the content and composition of chlorophylls (spectrophotometry), the content of soluble proteins (spectrophotometry), ionic permeability of cell membranes (conductometry), as well as the expression of various groups of genes associated with senescence (transcription factors that initiate senescence; genes controlling chlorophyll degradation; genes providing nitrogen remobilization and redistribution of catabolic products) were assessed. Expression studies were carried out using PCR after reverse transcription. The above parameters were also assessed against the background of treating plants with abscisic acid (the main hormonal regulator of senescence in plants).
Results: The initial chlorophyll content does not differ in Col-0 and *gdh1gdh2* mutant. But the chlorophyll degradation during seven days in the dark is much slower in the mutant plants. Chlorophyll *b* degradation is slower in the mutant too. Chlorophyll *a/b* ratio in the *gdh1gdh2* plant rosettes decreased within seven days in the dark due to a slower degradation of chlorophyll *b*. In the Col-0 rosettes, no significant change in the chlorophyll *a/b* ratio was observed.
Ion leakage conductometry test demonstrate cell membrane integrity and plant viability at all. The higher ion leakage the lower membrane integrity and plant viability. Plant viability decreased during the dark exposition for both genotypes, but much more quickly for *gdh1gdh2*. It is a sign of the type D stay-green phenotype – toxic metabolite accumulation.
To study the dynamics of changes in the gene expression during the normal course of dark-induced senescence in wild-type plants, we sampled the leaves based on their color: yellow (the lowest true ones), yellow-green (middle), and green (apical). Five chlorophyll degradation genes showed a significant increase in transcript levels from the youngest (green) to the oldest (completely yellow) leaves. To compare wild-type and stay-green *gdh1gdh2* plants, we selected only the first (the oldest) pair of true leaves from each rosette. Col-0 leaves demonstrate a gradual increase in expression of the chlorophyll degradation genes throughout the experiment. In the mutant plants, expression of all the studied chlorophyll degradation genes also had a tendency to increase after two days in the dark, but no continued increase in expression was observed later.
Errors in hormonal regulation are the often reason for senescence disruptions, so we examined plant sensitivity to abscisic acid (ABA) as the main and very strong senescence inducer. ABA sensitivity was not broken in the seeds of *gdh1gdh2*. But in the adult plants, we can see senescence induction in Col-0 and death without senescence signs in *gdh1gdh2*. Abscisic acid induced senescence is disrupted too. We examined *SAGs* (Senescence Associated Genes) expression in the upper and the lower leaves of our plants after four days in the dark. In the most cases there was no *SAGs* induction in the mutant plants. ABA treatment creates a deeper stage of senescence than four days of the dark. Now we can see some answer of *SAG113*, *SAG29*, *GLN1.1* in the mutant plants, but as a whole *SAGs* expression in the *gdh1gdh2* is not the same as in wild-type plants. Thus, senescence program in *gdh1gdh2* is not totally insensitive to ABA, but disrupted significantly.

Conclusion: The discovered facts suggest the existence of a functional connection between expression of the NAD(H)-dependent glutamate dehydrogenase genes and regulation of senescence processes in plants. The lack of glutamate dehydrogenase activity during prolonged dark exposition is probably the reason of disrupted senescence in *gdh1gdh2* plants. It may lead to metabolic switches followed by some toxic metabolites appearance. Thus, glutamate dehydrogenase is one of essential senescence regulators in Arabidopsis.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation, project number FWSS-2022-0005.

Список литературы/References

1. Tomlinson P.B. Chance and design in the construction of plants. In: Axioms and Principles of Plant Construction. Dortrecht: Springer, 1982. doi 10.1007/978-94-009-7636-8_9
2. Antonova I.S., Lagunova N.G. On modular organization of some groups of plants. *Zhurnal Obshchei Biologii.* 1999;60(1):58-59
3. Gan S. Concepts and Types of Senescence in Plants. *Methods Mol Biol.* 2018;1744:3-8. doi 10.1007/978-1-4939-7672-0_1
4. Garnik E.Yu., Belkov V.I., Tarasenko V.I., Rudikovskii A.V., Konstantinov Yu.M. Expression of glutamate dehydrogenase genes in *Arabidopsis thaliana* depends on the redox state of plastoquinone pool. *Plant Cell Tissue Organ Cult.* 2021;147:107-116. doi 10.1007/s11240-021-02111-5

doi 10.18699/bgrs2024-13-05

Сочетания вариантов митохондриальной ДНК как фактор риска развития нарушений сердечного ритма

Голубенко М.[1]*, Корепанов В.[2], Бабушкина Н.[1], Васильева О.[1], Валиахметов Н.[1], Зарубин А.[1], Атабеков Т.[2], Витт К.[2], Афанасьев С.[2], Назаренко М.[1]

[1] *НИИ медицинской генетики Томского НИМЦ, Томск, Россия*

[2] *НИИ кардиологии Томского НИМЦ, Томск, Россия*

* *maria.golubenko@medgenetics.ru*

Ключевые слова: митохондриальная ДНК; генетический полиморфизм; внезапная сердечная смерть; аритмии

Мотивация и цель: Одним из опасных осложнений ишемической болезни сердца (ИБС) являются нарушения ритма сердца, которые могут стать причиной внезапной сердечной смерти. Выявление факторов риска развития этих нарушений, в том числе генетических, является актуальной задачей. Вероятным звеном физиологических нарушений, развивающихся с возрастом и ведущих к повышению риска аритмии, являются митохондрии. Митохондриальная дисфункция может быть инициировать нарушения ритма сердца как напрямую, поскольку митохондрии производят необходимый для работы сердца АТФ, так и опосредованно, через повышение концентрации активных форм кислорода, повреждающих клеточные структуры, нарушение ионного баланса, реполяризацию мембран и апоптоз [1]. Митохондрии имеют собственный геном – митохондриальную ДНК (мтДНК), которая кодирует несколько субъединиц комплексов дыхательной цепи. Принимая во внимание, что мтДНК отличается высоким уровнем мутаций и, следовательно, популяционного полиморфизма, а также наследуется без рекомбинации, следует ожидать, что отдельные «нормальные» варианты мтДНК или их сочетания могут быть ассоциированы с изменением функциональных параметров митохондрий. Так как речь идет о популяционном полиморфизме, то величина этих эффектов мала, и они могут проявиться только в позднем возрасте, по мере развития ассоциированных с возрастом заболеваний.

Значимость вариаций мтДНК для фенотипа показана в ряде исследований, проведенных на цибридных линиях клеток, в которых отдельные генотипы мтДНК (гаплогруппы) различались по параметрам клеточного дыхания [2-5]. Кроме того, выявлены ассоциации полиморфизма мтДНК с различными многофакторными заболеваниями и их осложнениями, в том числе с внезапной смертью [6-10]. Цель исследования – поиск вариантов мтДНК и/или их сочетаний, влияющих на риск развития жизнеугрожающих аритмий на фоне ИБС.

Методы и алгоритмы: Исследована полная последовательность мтДНК в выборке пациентов с нарушениями ритма сердца (НРС), которым была проведена по показаниям имплантация кардиовертера-дефибриллятора (N=81, средний возраст 64 года) и в контрольной группе индивидов без признаков сердечно-сосудистых заболеваний (N=59, средний возраст 65 лет). Секвенирование мтДНК проводили с помощью метода NGS на приборе MiSeq (Illumina, США), процесс подготовки образцов и обработки данных NGS для мтДНК описан ранее [7]. Для каждого

образца замены аминокислот в митохондриальных генах белков были сгруппированы по комплексам дыхательной цепи митохондрий (I, III, IV, V), а нуклеотидные замены в генах РНК рассматривали отдельно для 12S рРНК, 16S РНК и транспортных РНК. Замены в «корне» основных гаплогрупп не учитывали (например, не была включена в анализ замена A12308G в гене тРНК лейцина, определяющая гаплогруппу U, замена аминокислоты Y304H в гене ND1 (T4216C, гаплогруппы T и J) и др.). Сравнение частот отдельных вариантов, гаплогрупп, сочетаний полиморфизмов проводили с помощью критерия Хи-квадрат или точного критерия Фишера.

Результаты: Сравнение частот основных гаплогрупп мтДНК в исследуемых выборках не выявило статистически значимых различий. Также не было выявлено различий по частоте отдельных замен, встречающихся в разных гаплогруппах, например, A10398G (T114A ND3), G3010A (16S rRNA). В то же время сравнение числа замен, зарегистрированных в разных группах генов в одном гаплотипе, показало преобладание таких «множественных» замен у пациентов с нарушениями ритма сердца, по сравнению с контрольной группой (Рисунок 1). Статистически значимые различия были получены при сравнении доли пациентов в выборке, имеющих замены аминокислот (для субъединиц, кодируемых мтДНК) в двух или трех комплексах дыхательной цепи ($p=0,003$; $X^2=10,37$), а также при сравнении доли пациентов, имеющих нуклеотидные замены одновременно в 12S рРНК и 16S рРНК ($p=0,03$; $X^2=4,7$).

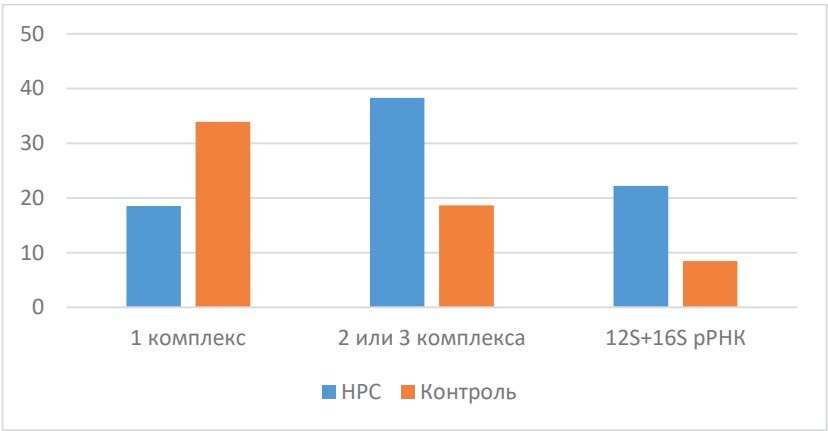


Рис. 1. Частота регистрации миссенс-замен одновременно в двух и более комплексах дыхательной цепи, а также одновременно в двух генах рРНК, кодируемых мтДНК, у пациентов с нарушениями ритма сердца и в контрольной выборке. Частоты указаны в процентах

Выводы: Таким образом, «нагруженность» митохондриального генома заменами в генах различных комплексов дыхательной цепи, а также одновременно в обоих генах рибосомных РНК, может иметь значение как фактор риска развития жизнеугрожающих аритмий у пациентов с ишемической болезнью сердца. Полученные результаты показывают, что присутствие в митохондриальном гаплотипе индивида множественных вариантов в различных генах, по-видимому, вносит вклад в усиление дисфункции митохондрий при возраст-ассоциированных заболеваниях и/или при старении.

Финансирование: Исследование поддержано Российским научным фондом (№ 24-24-00527).

Mitochondrial DNA variants combinations as a risk factor for developing of arrhythmia

Golubenko M.[1]*, Korepanov V.[2], Babushkina N.[1], Vasilyeva O.[1], Valiakhmetov N.[1], Zarubin A.[1], Atabekov T.[2], Vitt K.[2], Afanasiev S.[2], Nazarenko M.[1]

[1] *Research Institute of Medical Genetics, Tomsk NRMC, Tomsk, Russia*

[2] *Cardiology Research Institute, Tomsk NRMC, Tomsk, Russia*

* *maria.golubenko@medgenetics.ru*

Key words: mitochondrial DNA; genetic polymorphism; sudden cardiac death; arrhythmias

Motivation and Aim: One of the dangerous complications of coronary heart disease (CHD) is heart rhythm disturbances, which can cause sudden cardiac death. Identification of risk factors for the development of these conditions, including genetic ones, is an actual task. Mitochondria are a likely link in physiological dysregulation that develop with age and lead to an increased risk of arrhythmia. Mitochondrial dysfunction can be associated with cardiac arrhythmias either directly, as mitochondria produce the ATP for heart functioning, or indirectly, through an increase in the concentration of reactive oxygen species that damage cellular structures, disruption of the ion balance, membrane repolarization and apoptosis [1]. Mitochondria contain their own genome, mitochondrial DNA (mtDNA), which encodes several subunits of the respiratory chain complexes. Taking into account that human mtDNA is characterized by a high mutation rate and, therefore, extensive population polymorphism, and is inherited without recombination, it should be expected that "normal" mtDNA variants or their combinations may be associated with changes in the functional parameters of mitochondria. It should be noted that since we are talking about population polymorphism, the effect sizes for these variants is small, and they can only appear at a later age, as age-associated diseases develop.

The significance of mtDNA variations for phenotype is shown in a number of studies conducted on cybrid cell lines, in which individual mtDNA genotypes (haplogroups) differed in parameters of mitochondrial respiration [2-5]. In addition, associations of mtDNA polymorphism with various multifactorial diseases and their complications, including sudden death, have been identified [6-10]. The purpose of the study is to search for mtDNA variants and/or their combinations that affect the risk of developing life-threatening arrhythmias as a complication of coronary artery disease.

Methods and Algorithms: The complete sequence of mtDNA was studied in a sample of patients with cardiac arrhythmias, who underwent implantation of a cardioverter-defibrillator (N=81, average age 64 years) and in a control group of individuals without signs of cardiovascular diseases (N=59, average age 65 years). MtDNA sequencing was carried out using the NGS method on a MiSeq instrument (Illumina, USA), the process of sample preparation and processing of NGS data for mtDNA is described in [7]. For each sample, amino acid substitutions were registered in the mitochondrial genes of the subunits of the mitochondrial respiratory chain complexes, depending on the complex (I, III, IV, V), as well as nucleotide substitutions in the ribosomal (12S and 16S) and

transport RNA genes encoded by the mitochondrial genome. Substitutions in the "root" of the main haplogroups were not taken into account (for example, the A12308G substitution in the leucine tRNA gene, which determines haplogroup U, the Y304H amino acid substitution in the ND1 gene (T4216C), etc., were not included in the analysis). Comparison of the frequencies of individual variants, haplogroups, and combinations of polymorphisms was carried out using the Chi-square test or Fisher's exact test.

Results: Comparison of the main mtDNA haplogroups frequencies in the studied samples did not reveal statistically significant differences. There were also no differences in the frequency of individual substitutions repeatedly found in different haplogroups, for example, A10398G (T114A in ND3), G3010A (16S rRNA). At the same time, a comparison of the number of substitutions registered in different "classes" of genes in one haplotype showed the predominance of such "multiple" substitutions in patients with cardiac arrhythmias compared to the control group (Figure 1). Statistically significant differences were obtained when comparing the proportion of patients with amino acid substitutions in two or three complexes of the respiratory chain ($p=0.003$; $X^2=10.37$), as well as when comparing the proportion of patients with nucleotide substitutions simultaneously in 12S rRNA and 16S rRNA ($p=0.03$; $X^2=4.7$).

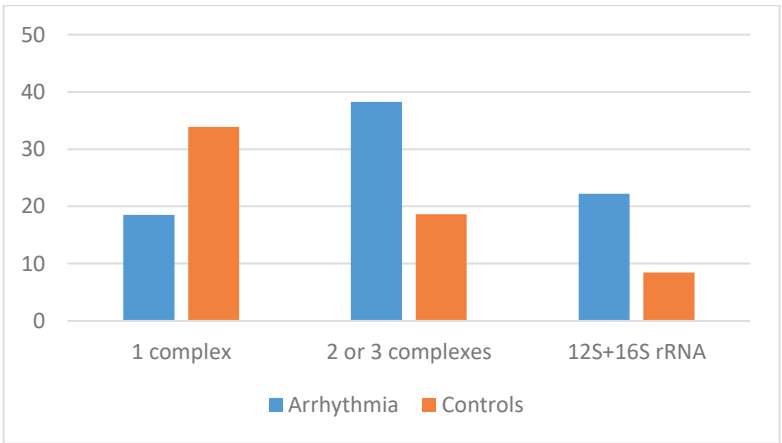


Fig. 1. Frequency of missense substitutions in two or more complexes of the respiratory chain simultaneously, as well as in both rRNA genes encoded by mtDNA, in patients with cardiac arrhythmias and in the control sample. Frequencies are given as percentages

Conclusion: Thus, the "variation load" of the mitochondrial genome with substitutions in the genes of different respiratory chain complexes, as well as simultaneously in both ribosomal RNA genes, may have impact as a risk factor for the development of life-threatening arrhythmias in patients with coronary heart disease. The results indicate that the presence of multiple variants in different genes within an individual's mitochondrial haplotype appears to contribute to increased mitochondrial dysfunction in age-associated diseases and/or aging.

Funding: The study is supported by Russian Science Foundation (No. 24-24-00527).

Список литературы/References

1. Yang K.C., Bonini M.G., Dudley S.C. Jr. Mitochondria and arrhythmias. *Free Radic Biol Med*. 2014;71:351-361
2. Lin T.K. et al. The creation of cybrids harboring mitochondrial haplogroups in the Taiwanese population of ethnic Chinese background: an extensive in vitro tool for the study of mitochondrial genomic variations. *Oxid Med Cell Longev*. 2012;2012:824275
3. Kenney M.C. et al. Molecular and bioenergetic differences between cells with African versus European inherited mitochondrial DNA haplogroups: implications for population susceptibility to diseases. *Biochim Biophys Acta.* 2014;1842(2):208-219
4. Zhou H. et al. Generation and Bioenergetic Profiles of Cybrids with East Asian mtDNA Haplogroups. *Oxid Med Cell Longev*. 2017;2017:1062314
5. Panvini A.R. et al. Differential mitochondrial and cellular responses between H vs. J mtDNA haplogroup-containing human RPE transmitochondrial cybrid cells. *Exp Eye Res.* 2022;219:109013
6. Голубенко М.В. и др. Ассоциации полиморфизма митохондриальной ДНК с инфарктом миокарда и прогностически значимыми признаками атеросклероза. *Молекулярная биология*. 2015;49(6):968-976
   [Golubenko M.V. et al. Association of mitochondrial DNA polymorphism with myocardial infarction and prognostic signs for atherosclerosis. *Mol Biol.* 2015;49(6):968-976 (in Russian)]
7. Голубенко М.В. и др. Ассоциация вариантов гаплогруппы H1 митохондриальной ДНК с риском сердечно-сосудистых катастроф. *Научные результаты биомедицинских исследований*. 2019;5(4):19-31
   [Golubenko M.V. et al. Association of the mitochondrial DNA haplogroup H1 variants with the risk of acute cardiovascular events. *Res Results Biomed*. 2019;5(4):19-31 (in Russian)]
8. Голубенко М.В. и др. Полиморфизм митохондриальной ДНК и ишемия миокарда: ассоциация гаплогруппы H. *Сибирский журнал клинической и экспериментальной медицины*. 2021;36(4):70-77
   [Golubenko M.V. et al. Mitochondrial DNA polymorphism and myocardial ischemia: Association of haplogroup H with heart failure. *Sib J Clin Exp Med*. 2021;36(4):70-77 (in Russian)]
9. Kytövuori L. et al. Mitochondrial DNA variation in sudden cardiac death: a population-based study. *Int J Legal Med*. 2020;134(1):39-44
10. Roselló-Díez E. et al. Mitochondrial genetic effect on atrial fibrillation: A case-control study. *Mitochondrion.* 2021;56:15-24

doi 10.18699/bgrs2024-13-06

Влияние делеций генов биосинтеза сероводорода *CBS*, *CSE* и *MST* при ограниченном потреблении серосодержащих аминокислот (метионина и цистеина) на жизнеспособность *Drosophila melanogaster*

Горбунова А.А.*, Шапошников М.В., Бабак Т.В., Коваль Л.А., Москалев А.А.
Институт биологии ФИЦ Коми НЦ УрО РАН, Сыктывкар, Республика Коми, Россия
** gorbunova.a@ib.komisc.ru*

Ключевые слова: сероводород; делеции; цистатионин-β-синтаза; цистатионин-γ-лиаза; 3-меркаптопируват-сульфуртрансфераза; метионин; цистеин

Мотивация и цель: Эндогенный синтез сероводорода (H_2S) – биологический процесс, происходящий внутри клеток и тканей организма, который включает серию ферментативных реакций по пути транссульфурации. Этот путь отвечает за метаболизм серосодержащих аминокислот, таких как метионин и цистеин. Цистатионин-β-синтаза (CBS) – один из основных ферментов пути транссульфурации. Он катализирует превращение гомоцистеина и серина, производных метионина, в цистатионин. На следующем этапе транссульфурации цистатионин-γ-лиаза (CSE) катализирует распад цистатионина на цистеин и α-кетобутират. Затем подключается 3-меркаптопируватсульфуратрансфераза (3MST), работающий совместно с цистеинаминотрансферазой (CAT). CAT сначала превращает цистеин и α-кетоглутарат в глутамат и 3-меркаптопируват. Затем 3MCT катализирует перенос атома серы от 3-меркаптопирувата к подходящему акцептору, что приводит к образованию H_2S.
Целью настоящего исследования было изучение влияния делеции генов *CBS, CSE* и *MST* в различных сочетаниях на продолжительность жизни, стрессоустойчивость и двигательную активность *D. melanogaster* при контроле содержания метионина и цистеина в их рационе.
Фон: Ранее в работах нашего коллектива [1] и в совместных исследованиях с коллегами из Института молекулярной Биологии РАН им. Энгельгардта было показано, что усиленная экспрессия генов *CBS* и *CSE,* а также делеция этих генов вела к увеличению продолжительности жизни, но при этом контроля за содержанием метионина и цистеина в составе диеты не было [2].
Методы: В работе использовали лабораторные линии *D. melanogaster* из коллекции ИМБ РАН им. Энгельгардта, гомозиготные по делециям генов *CBS* (обозначена как *CBS-/-*) и *3-MST* (*3-MST-/-*), а также гибридная линия с делециями двух генов *CBS* и *CSE* (*CBS-/-; CSE-/-*). В качестве контроля использовалась линия *58492* (BDSC_58492, Блумингтон, США, обозначена как (*CBS+/+, CSE+/+*)).
Контрольных и экспериментальных мух содержали на синтетической питательной среде, без дрожжей, с точно контролируемым составом углеводов, аминокислот, липидов, витаминов и микроэлементов (смесь TD.210411, ENVIGO, США), отличающейся содержанием серосодержащих аминокислот метионина (Met) и цистеина (Cys): контрольная среда, содержащая обе аминокислоты (Met+, Cys+);

среда, содержащая только метионин (Met+, Cys-); среда, содержащая только цистеин (Met-, Cys+) и среда, с дефицитом обеих аминокислот (Met-, Cys-).
Для оценки продолжительности жизни каждый день подсчитывали количество умерших мух. Замену питательной среды на свежую проводили три раза в неделю. Для индукции окислительного стресса двухнедельных мух содержали на питательной среде с добавлением 20 mM параквата (Merk, Германия). При оценке устойчивости к гипертермии, двухнедельных мух содержали при температуре 35 °C. Обработку паракватом и воздействие гипертермии проводили до гибели всех особей. Возрастные изменения подвижности исследовали путем оценки спонтанной двигательной активности мух с помощью монитора локомоторной активности LAM25H (Trikinetics, США) один раз в неделю.
Для определения статистической значимости различий в медианной и максимальной продолжительности жизни использовали точный критерий Фишера. Лог-ранговый тест использовался для оценки статистической значимости различий между кривыми выживаемости. Двусторонний дисперсионный анализ (ANOVA) вместе с апостериорным тестом Тьюки HSD был проведен для установления статистической значимости отклонений в двигательной активности.
Результаты: При анализе продолжительности жизни линий *D. melanogaster* с делециями по исследуемым генам при различном содержании аминокислот в питательной среде, было отмечено, что продолжительность жизни контрольной линии (*CBS+/+, CSE+/+*) и линии *3-MST-/-* была выше при содержании на среде с добавлением метионина (Met+, Cys+) и (Met+, Cys-) по сравнению с вариантами сред с дефицитом метионина (Met-, Cys+) и (Met-, Cys-). Было обнаружено, что отсутствие метионина, резко сокращает продолжительность жизни мух. Для особей линии с двойной делецией (*CBS-/-*; *CSE-/-*) наибольшая продолжительность жизни наблюдалась при содержании на контрольном варианте среды (Met+, Cys+), за ним следовал вариант с цистеиновой средой (Met-, Cys+). Наименьшую продолжительность жизни имели варианты, содержащиеся на средах с дефицитом обеих аминокислот (Met-, Cys-) и с добавлением метионина, но с дефицитом цистеина (Met+, Cys-). Последнее наблюдение можно объяснить накоплением гомоцистеина, токсичного промежуточного продукта метаболизма в пути транссульфурации, что может приводить к различным физиологическим нарушениям. Кроме того, у особей линии *CBS-/-* наибольшая продолжительность жизни наблюдалась в вариантах, содержащихся на питательной среде с цистеином, независимо от добавки метионина (Met-, Cys+) и (Met+, Cys+). Наименьшую продолжительность жизни имели мухи в варианте с добавлением метионина, но с дефицитом цистеина (Met+, Cys-).
В условиях окислительного стресса самки контрольной линии (*CBS+/+, CSE+/+*) продемонстрировали лучшую выживаемость на среде с дефицитом метионина (Met-, Cys+), в то время как самцы имели лучшую выживаемость на среде с добавлением метионина (Met+, Cys-). Особи обоих полов линии *CBS-/-* имели лучшую выживаемость на контрольной среде (Met+, Cys+). У линии с двойной делецией (*CBS-/-*; *CSE-/-*) контрольная среда (Met+, Cys+) и среда с дефицитом метионина (Met-, Cys+) обеспечивали наилучшие условия выживания. У линии *3-MST-/-* самцы имели лучшую выживаемость на контрольной среде (Met+, Cys+) и среде с добавлением метионина и дефицитом цистеина (Met+,

Cys-), тогда как самки имели лучшую выживаемость при добавлении какой-либо аминокислоты.
При изучении воздействия гипертермии было отмечено, что линия *CBS-/-* имела самую высокую выживаемость на питательной среде с дефицитом метионина (Met-, Cys+).
Также установлено, что двигательная активность *D. melanogaster* существенно варьирует в зависимости от наличия или отсутствия аминокислот, их сочетаний, возраста и пола. В частности, влияние метионина на двигательную активность различалось у самцов и самок контрольной линии (*CBS+/+, CSE+/+*). У самок добавление в рацион метионина снижало двигательную активность. У самцов, наоборот, присутствие метионина в среде ее повышало.
Выводы: Согласно полученным результатам, на продолжительность жизни, выживаемость и двигательную активность *D. melanogaster* влияет взаимодействие между ферментами, продуцирующими сероводород (CBS, CSE и 3MST) и пищевыми серосодержащими аминокислотами (метионином и цистеином).
Присутствие метионина и цистеина в рационе *D. melanogaster* оказывает существенное влияние на продолжительность жизни. В частности, было обнаружено, что отсутствие метионина резко сокращает продолжительность жизни контрольной линии мух (*CBS+/+, CSE+/+*) и линии *3-MST-/-*. Результаты также показывают, что линия с двойной делецией (*CBS-/-; CSE-/-*) и линия с одиночной делецией (*CBS-/-*) по-разному реагируют на присутствие или отсутствие метионина и цистеина в их рационе. У линии с двойной делецией вариант контрольной среды (Met+, Cys+) имел наибольшую продолжительность жизни, за ним следовал вариант с добавлением цистеина (Met-, Cys+). Это говорит о том, что наличие цистеина более важно для долголетия мух на этом генетическом фоне, чем присутствие метионина. Наименьшую продолжительность жизни имели варианты с дефицитом обеих аминокислот (Met-, Cys-) и вариант с добавлением метионина (Met+, Cys-), что могло быть связано с накоплением гомоцистеина. В целом, исследование подчеркивает важность сбалансированной диеты, включающей как метионин, так и цистеин, для долголетия *D. melanogaster* и предполагает, что генетический фон мух может влиять на их реакцию на различные диетические условия.
Оптимальные условия для выживания при окислительном стрессе и гипертермии различались, у разных линий и полов, что указывает на то, что влияние различных факторов на выживаемость является сложным и может зависеть от генетического фона и факторов, специфичных для пола. Это подчеркивает важность рассмотрения пола как биологической переменной в исследованиях, изучающих влияние пищевых компонентов на физиологические и поведенческие результаты. Необходимы дальнейшие исследования, чтобы полностью понять основные механизмы и потенциальное применение этих результатов.
Финансирование: Работа выполнена в рамках гранта РНФ № 17-74-30030.

Effect of deletions of the *CBS, CSE* and *MST* hydrogen sulfide biosynthesis genes under restricted intake of sulfur-containing amino acids (methionine and cysteine) on the viability of *Drosophila melanogaster*

Gorbunova A.A.*, Shaposhnikov M.V., Babak T.V., Koval L.A., Moskalev A.A.
Institute of Biology, Federal Research Center Komi Scientific Center, Ural Branch of RAS, Syktyvkar, Komi Republic, Russia
** gorbunova.a@ib.komisc.ru*

Key words: hydrogen sulfide; deletions; cystathionine-β-synthase; cystathionine-γ-lyase; 3-mercaptopyruvate sulfurtransferase; methionine; cysteine

Motivation and Aim: Endogenous hydrogen sulfide (H_2S) synthesis is a biological process occurring within the cells and tissues of the body that involves a series of enzymatic reactions via the transsulfuration pathway. This pathway is responsible for the metabolism of sulfur-containing amino acids such as methionine and cysteine. Cystathionine-β-synthase (CBS) is one of the major enzymes of the transsulfuration pathway. It catalyzes the conversion of homocysteine and serine derivatives of methionine to cystathionine. In the next step of transsulfuration, cystathionine-γ-lyase (CSE) catalyzes the breakdown of cystathionine into cysteine and α-ketobutyrate. Then 3-mercaptopyruvate sulfurathransferase (3MST), working in conjunction with cysteine aminotransferase (CAT), is involved. CAT first converts cysteine and α-ketoglutarate to glutamate and 3-mercaptopyruvate. 3MCT then catalyzes the transfer of a sulfur atom from 3-mercaptopyruvate to a suitable acceptor, resulting in the formation of H_2S.
The aim of the present study was to investigate the effect of deletion of *CBS, CSE* and *MST* genes in different combinations on longevity, stress tolerance and locomotor activity of *D. melanogaster* while controlling methionine and cysteine content in their diet.
Background: Earlier in the works of our team [1] and in joint studies with colleagues from the Engelhardt Institute of Molecular Biology of the Russian Academy of Sciences, it was shown that enhanced expression of *CBS* and *CSE* genes, as well as deletion of these genes, led to an increase in life span, but at the same time the content of methionine and cysteine in the diet was not controlled experimentally [2].
Methods and Algorithms: Laboratory lines of *D. melanogaster* from the Engelhardt IMB RAS collection homozygous for deletions of CBS (designated as *CBS-/-*) and 3-MST (*3-MST-/-*) genes, as well as a hybrid line with deletions of two genes *CBS* and *CSE* (*CBS-/-; CSE-/-*). Line 58492 (BDSC_58492, Bloomington, USA, designated as (*CBS+/+, CSE+/+*)) was used as control.
Control and experimental flies were maintained on synthetic nutrient medium, without yeast, with precisely controlled composition of carbohydrates, amino acids, lipids, vitamins and trace elements (mix TD.210411, ENVIGO, USA), differing in the content of the sulfur-containing amino acids methionine (Met) and cysteine (Cys): control medium containing both amino acids (Met+, Cys+); medium containing only methionine (Met+, Cys-); medium containing only cysteine (Met-, Cys+) and medium deficient in both amino acids (Met-, Cys-).

The number of flies that died was counted each day to assess longevity. The nutrient medium was replaced with fresh medium three times a week. For induction of oxidative stress, two-week-old flies were maintained on nutrient medium supplemented with 20 mM paraquat (Merk, Germany). For evaluation of resistance to hyperthermia, two-week-old flies were kept at 35 °C. Paraquat treatment and exposure to hyperthermia were carried out until all individuals were killed. Age-related changes in motility were investigated by assessing the spontaneous locomotor activity of flies using a LAM25H locomotor activity monitor (Trikinetics, USA) once a week.

Fisher's exact test was used to determine the statistical significance of differences in median and maximum life expectancy. Log-rank test was used to assess the statistical significance of differences between survival curves. Two-way analysis of variance (ANOVA) along with Tukey's HSD posterior test was performed to establish the statistical significance of variance in motor activity.

Results: When analyzing the life span of *D. melanogaster* lines with deletions in the studied genes at different content of amino acids in the nutrient medium, it was noted that the life span of the control line (*CBS*+/+, *CSE*+/+) and the *3-MST*-/- line was higher when the medium was supplemented with methionine (Met+, Cys+) and (Met+, Cys-) compared to the variants of the medium with methionine deficiency (Met-, Cys+) and (Met-, Cys-). Lack of methionine was found to drastically reduce the lifespan of flies. For individuals of the line with double deletion (*CBS-/-; CSE-/-*), the longest life span was observed when kept on control medium (Met+, Cys+), followed by the variant with cysteine medium (Met-, Cys+). The shortest life span was observed for variants kept on media deficient in both amino acids (Met-, Cys-) and with the addition of methionine but deficient in cysteine (Met+, Cys-). The latter observation could be explained by the accumulation of homocysteine, a toxic metabolic intermediate in the transsulfuration pathway, which may lead to various physiological disorders. In addition, in individuals of the *CBS-/-* line, the longest lifespan was observed in variants maintained on nutrient medium with cysteine, regardless of methionine supplementation (Met-, Cys+) and (Met+, Cys+). Flies in the methionine supplemented but cysteine deficient (Met+, Cys-) variant had the shortest lifespan.

Under oxidative stress conditions, control females (*CBS*+/+, *CSE*+/+) showed better survival on methionine-deficient medium (Met-, Cys+), while males had better survival on methionine-supplemented medium (Met+, Cys-). Individuals of both sexes of the *CBS-/-* line had better survival on control medium (Met+, Cys+). In the double deletion line (*CBS-/-; CSE-/-*), control medium (Met+, Cys+) and methionine-deficient medium (Met-, Cys+) provided the best survival conditions. In the *3-MST-/-* line, males had the best survival on control medium (Met+, Cys+) and methionine- and cysteine-deficient medium (Met+, Cys-), whereas females had the best survival when any amino acid was added.

When the effects of hyperthermia were studied, it was observed that *CBS-/-* line had the highest survival rate on nutrient medium with methionine deficiency (Met-, Cys+).

The locomotor activity of *D. melanogaster* was also found to vary significantly depending on the presence or absence of amino acids, their combinations, age and sex. In particular, the effect of methionine on locomotor activity differed between males and control line females (*CBS*+/+, *CSE*+/+). In females, the addition of methionine to the diet decreased locomotor activity. In males, on the contrary, the presence of methionine in the medium increased it.

Conclusions: According to the results obtained, the longevity, survival and locomotor activity of *D. melanogaster* are influenced by the interaction between hydrogen sulfide-producing enzymes (CBS, CSE and 3MST) and dietary sulfur-containing amino acids (methionine and cysteine).

The presence of methionine and cysteine in the diet of *D. melanogaster* has a significant effect on longevity. In particular, the absence of methionine was found to dramatically reduce the lifespan of the control fly line (*CBS+/+, CSE+/+*) and the *3-MST-/-* line. The results also show that a line with a double deletion (CBS-/-; CSE-/-) and a line with a single deletion (CBS-/-) respond differently to the presence or absence of methionine and cysteine in their diet. In the double deletion line, the control medium variant (Met+, Cys+) had the highest longevity, followed by the cysteine supplemented variant (Met-, Cys+). This suggests that the presence of cysteine is more important for fly longevity in this genetic background than the presence of methionine. The variants deficient in both amino acids (Met-, Cys-) and the variant with methionine supplementation (Met+, Cys-) had the lowest longevity, which could be due to homocysteine accumulation. Overall, the study emphasizes the importance of a balanced diet including both methionine and cysteine for the longevity of *D. melanogaster* and suggests that the genetic background of flies may influence their response to different dietary conditions.

Optimal conditions for survival under oxidative stress and hyperthermia differed, across lines and sexes, indicating that the effects of various factors on survival are complex and may depend on genetic background and sex-specific factors. This highlights the importance of considering sex as a biological variable in studies examining the effects of nutritional components on physiological and behavioral outcomes.

Further research is needed to fully understand the underlying mechanisms and potential applications of these findings.

Funding: The work was carried out within the framework of the Russian Science Foundation grant No. 17-74-30030.

Список литературы/References

1. Shaposhnikov M. et al. Overexpression of *CBS* and *CSE* genes affects lifespan, stress resistance and locomotor activity in *Drosophila melanogaster*. *Aging (Albany NY).* 2018;10(11):3260-3272. doi 10.18632/aging.101630
2. Shaposhnikov M.V. et al. Deletions of the *cystathionine-β-synthase (CBS)* and cystathionine-*γ*-*lyase (CSE)* genes, involved in the control of hydrogen sulfide biosynthesis, significantly affect lifespan and fitness components of *Drosophila melanogaster*. *Mech Ageing Dev*. 2022;203:111656. doi 10.1016/j.mad.2022.111656

doi 10.18699/bgrs2024-13-07

Изучение молекулярных механизмов изменения транскриптомного возраста в период раннего эмбриогенеза мышей на уровне генной экспрессии одиночных клеток

Давитадзе М.С.[1]*, Дмитриев С.Е.[2], Тышковский А.Э.[2, 3]

[1] *Факультет биоинженерии и биоинформатики, Московский государственный университет, Москва, Россия*
[2] *НИИ физико-химической биологии имени А.Н. Белозерского, Московский государственный университет, Москва, Россия*
[3] *Brigham and Women's Hospital, Медицинская Школа Гарварда, Бостон, США*
* *maria.davitadze@mail.ru*

Ключевые слова: часы старения; омоложение; эмбриогенез; scRNA-seq; траектории дифференцировки

Мотивация и цель: Современные исследования показывают, что половые клетки подвержены возрастным изменениям: так, в сперматозоидах накапливаются повреждения, ассоциированные со старением, такие как эпигенетические изменения и мутации [1], а эпигенетический возраст гамет и зиготы сопоставим с возрастом соматических клеток того же организма [2]. В связи с этим была выдвинута гипотеза, что после оплодотворения, в процессе эмбриогенеза, должно происходить обращение возрастных изменений, то есть омоложение. В подтверждение этой гипотезы в недавних исследованиях с помощью эпигенетических часов старения было продемонстрировано, что после оплодотворения биологический возраст эмбрионов мышей действительно снижается [2]. При этом конкретные механизмы, лежащие в основе этого процесса, и их универсальность для всех эмбриональных клеточных популяций остаются практически неизученными. Целью нашей работы являлось более детальное изучение механизмов изменения биологического возраста в отдельных клеточных популяциях в процессе раннего эмбриогенеза мышей.
Методы и алгоритмы: Для анализа мы использовали публично доступные данные секвенирования 116312 одиночных клеток 350 эмбрионов мыши, собранных в девяти последовательных временных точках от 6.5 до 8.5 дней после оплодотворения [3]. Восстановление траекторий дифференцировки проводили с помощью метода оптимального транспорта Waddington-OT [4]. Биологический и хронологический возраст, а также ожидаемый риск смертности предсказывали с помощью мульти-тканевых транскриптомных часов старения, представляющих собой модель линейной регрессии, обученной предсказывать возраст на данных секвенирования РНК. Также мы использовали модульные транскриптомные часы, построенные на наборах ко-регулируемых генов, связанных с определенными функциональными процессами клеток.
Результаты: Мы проанализировали изменение суммарного биологического возраста эмбрионов, измеренного с помощью транскриптомных часов, в процессе гаструляции и раннего органогенеза. Мы обнаружили, что в этот период эмбрионального развития транскриптомный возраст падает, то есть профиль генной экспрессии меняется в сторону, противоположную изменениям, ассоциированным с возрастом и увеличением смертности. Эти результаты соотносятся с данными предыдущих исследований на уровне метилирования ДНК

[2]. Для того чтобы расширить понимание того, за счет чего наблюдается подобный эффект, мы проанализировали изменение транскриптомного возраста на уровне отдельных клеточных популяций. Для этого мы использовали данные scRNA-seq и с помощью метода оптимального транспорта восстановили траектории развития клеточных типов. Мы показали, что снижение биологического возраста происходит в ходе развития всех клеточных типов за исключением клеток первичной энтодермы, хотя скорость снижения молекулярного возраста варьируется между разными популяциями. Быстрее всего возраст падает у глоточной мезодермы, параксиальной мезодермы и эндотелия. С помощью модульных часов мы оценили, какие функциональные процессы вносят наибольший вклад в наблюдаемое снижение возраста. В целом молекулярные пути, отвечающие за падение биологического возраста, оказались схожи у различных клеточных популяций. Наиболее заметный эффект наблюдался для модулей, связанных с метаболизмом ксенобиотиков, коагуляцией, клеточной адгезией, клеточным циклом и модификацией хроматина. Интересно, что некоторые клеточные процессы, такие как метаболизм жирных кислот и ответ на тепловой стресс, напротив, показали увеличение возраста в процессе гаструляции и раннего органогенеза для большинства типов клеток, что может указывать на отсутствие унифицированной возрастной динамики клеточных процессов в ходе эмбриогенеза мышей.

Выводы: Транскриптомный возраст снижается в ходе развития большинства клеточных типов в период гаструляции и раннего органогенеза мышей (с 6.5 по 8.5 день), при этом у различных клеточных типов наблюдается разная скорость снижения возраста. Снижение транскриптомного возраста в разных клеточных популяциях происходит за счет схожих молекулярных путей, однако некоторые клеточные процессы начинают проявлять признаки старения уже на данной стадии эмбриогенеза, что может указывать на индивидуальный характер траектории омоложения и старения различных компонентов клеток в процессе эмбрионального развития млекопитающих.

Exploring the molecular mechanisms behind changes in transcriptomic age during early mouse embryogenesis at the level of single-cell gene expression

Davitadze M.S.[1]*, Dmitriev S.E.[2], Tyshkovskiy A.E.[2, 3]

[1] *Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia*
[2] *Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow, Russia*
[3] *Brigham and Women's Hospital, Harvard Medical School, Boston, USA*
** maria.davitadze@mail.ru*

Key words: aging clocks; rejuvenation; embryogenesis; scRNA-seq; differentiation trajectories

Motivation and Aim: Modern research shows that gametes are subject to age-related changes. For instance, sperm cells accumulate aging-associated damages, such as epigenetic changes and mutations [1], and the epigenetic age of gametes and zygotes matches that of somatic cells of the same organism [2]. This observation has led to the hypothesis that embryogenesis involves a reversal of age-related changes, essentially resulting in rejuvenation. Supporting this hypothesis, recent studies using epigenetic aging clocks have shown that the biological age of mouse embryos decreases after

fertilization [2]. Nevertheless, the specific mechanisms underlying this process and their universality across all embryonic cell populations are still poorly understood. The aim of our study was to further investigate the mechanisms of biological age change in individual cell populations during early mouse embryogenesis.

Methods and Algorithms: We analyzed publicly available sequencing data from 116312 single cells from 350 mouse embryos, collected at nine consecutive time points from 6.5 to 8.5 days post-fertilization [3]. We reconstructed differentiation trajectories using the Waddington-OT optimal transport method [4]. Biological and chronological age, along with expected mortality risk, were predicted using multi-tissue transcriptomic aging clocks – a linear regression model trained on RNA sequencing data. We also employed modular transcriptomic clocks based on sets of co-regulated genes linked to specific cellular functions.

Results: Our analysis of changes in the total biological age of embryos, as measured by transcriptomic clocks, during gastrulation and early organogenesis showed a decrease in transcriptomic age. This suggests that the gene expression profile shifts away from age-related changes and increased mortality. These findings are consistent with previous DNA methylation studies [2]. To further explore the mechanisms behind this effect, we analyzed changes in transcriptomic age at the level of individual cell populations using scRNA-seq data and the Waddington-OT method to reconstruct cell type development trajectories. We found that biological age decreases in all cell types except for the extraembryonic endoderm cells, although the rate of decrease in transcriptomic age varies among cell populations. The most substantial decreases occur in the pharyngeal mesoderm, paraxial mesoderm, and endothelium. By using modular clocks, we evaluated which functional processes contribute most significantly to age reduction. We found that, overall, the molecular pathways responsible for decreasing biological age are similar across cell populations. The most noticeable effect was observed in modules associated with xenobiotic metabolism, coagulation, cell adhesion, cell cycle, and chromatin modification. Interestingly, some cellular processes, such as fatty acid metabolism and the heat stress response, showed an increase in age during gastrulation and early organogenesis for most cell types, which may suggest a lack of uniform age dynamics in cellular processes during mouse embryogenesis.

Conclusion: Transcriptomic age decreases during the development of most cell types from day 6.5 to 8.5 of mouse embryogenesis. Various cell types experience different rates of age reduction. The reduction in transcriptomic age in different cell populations is driven by similar molecular pathways, however, some cellular processes begin to show signs of aging at this stage of embryogenesis, which may indicate the individual character of the rejuvenation and aging trajectories of various cell components during the embryonic development of mammals.

Список литературы/References

1. Jenkins T.G., Aston K.I., Pflueger C., Cairns B.R., Carrell D.T. Age-associated sperm DNA methylation alterations: possible implications in offspring disease susceptibility. *PLoS Genet*. 2014;10(7):e1004458. doi 10.1371/journal.pgen.1004458
2. Kerepesi C., Zhang B., Lee S.G., Trapp A., Gladyshev V.N. Epigenetic clocks reveal a rejuvenation event during embryogenesis followed by aging. *Sci Adv*. 2021;7(26):eabg6082. doi 10.1126/sciadv.abg6082
3. Pijuan-Sala B., Griffiths J.A., Guibentif C. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*. 2019;566(7745):490-495. doi 10.1038/s41586-019-0933-9
4. Schiebinger G., Shu J., Tabaka M. et al. Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell*. 2019;176(4):928-943.e22. doi 10.1016/j.cell.2019.01.006

doi 10.18699/bgrs2024-13-08

Однонуклеотидные полиморфизмы, ассоциированные с развитием фенотипа преждевременного старения у крыс OXYS

Девяткин В.*, Редина О., Колосова Н., Муралёва Н.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
** devyatkin@bionet.nsc.ru*

Ключевые слова: крысы OXYS; SNP; гипертензия; болезнь Альцгеймера; старение; возраст-зависимые заболевания

Мотивация и цель: Старение – основной фактор риска возраст-зависимых заболеваний, распространённость которых растет на фоне увеличения продолжительности жизни. Их развитие в более раннем возрасте является проявлением преждевременного старения, более позднее становится основой успешного старения, долголетия. Более 60 % людей в возрасте старше 65 лет страдают несколькими заболеваниями, «набор» и возраст манифестации которых определяют генетические и экологические факторы, качество жизни. Известно, что артериальная гипертония, которой страдает 7 % людей в возрасте до 40 лет и две трети – старше 60 лет, способствует развитию болезни Альцгеймера (БА) и возрастной макулярной дегенерации (ВМД) – неизлечимых нейродегенеративных заболеваний, которые зачастую развиваются одновременно. Генетические основы предрасположенности к развитию коморбидных возраст-зависимых состояний остаются не ясными. Продуктивным подходом к изучению их природы является использование биологических моделей, генетическая однородность и стандартные условия содержания которых снижают влияние внешних факторов и повышают воспроизводимость результатов. Целью настоящей работы был поиск нуклеотидных вариантов, с которыми связано раннее развитие сходного с гериатрическими заболеваниями человека фенотипа у крыс линии OXYS (ИЦиГ СО РАН) – уникальной модели преждевременного старения. На фоне умеренно повышенного кровяного давления у крыс OXYS развиваются катаракта, кардиомиопатия, ретинопатия, аналогичная ВМД у людей, и ключевые признаки БА при отсутствии мутаций в генах *App*, *Psen1* и *Psen2*, специфичных для наследственной формы заболевания.
Методы и алгоритмы: Анализировали данные RNA-seq, полученные ранее для образцов префронтальной коры, гиппокампа и сетчатки крыс OXYS. Риды картировали на сборку референсного генома Rnor 6.0. Из выявленных полиморфных вариантов с минимальным качеством картирования, равным 100, были выбраны позиции, генотип которых находился в гомозиготном состоянии по меньшей мере у трёх животных в одной экспериментальной группе. Глубина чтения (DP) должна быть ≥10 по крайней мере у одного животного.
Выявленные у крыс OXYS SNPs сравнивались с данными геномного и транскриптомного секвенирования 45 других линий крыс, сравнение проводили только по локусам, секвенированным в транскриптомном анализе крыс OXYS.

Результаты: Сравнение данных транскриптомов префронтальной коры, гиппокампа и сетчатки крыс OXYS с референсным геномом крыс BN/NHsdMcwi выявило 42478 SNPs в 9903 генах, из них 40373 SNP в 9699 генах были также выявлены в последовательностях геномов 11 линий/сублиний, использующихся в качестве нормотензивного контроля, и 22 линий/сублиний крыс с фенотипом, не связанным с артериальной гипертонией или старением. Среди оставшихся 2105 SNP проведен поиск генетических вариантов, которые могут влиять на проявление фенотипа преждевременного старения у крыс OXYS. Гены, несущие мутации у крыс OXYS, связаны с процессингом и презентацией антигенов, ростом клеток при развитии, положительной регуляцией локализации белка в клетке, регуляцией связывающих белков, процессами в многоклеточном организме, клеточным старением, главным комплексом гистосовместимости I и фагосомами. Выявлено семь нуклеотидных замен, способных оказывать значительный эффект на структуру транскрипта. Наиболее существенно укороченные белки ожидаются для *Rims2*, *Lemd2*, *Slc25a36l1* и *AABR07045405.1*, в первых трёх из них ожидается потеря функционально значимых доменов. В первом экзоне двух последних генов происходит сдвиг рамки считывания, что также может привести к накоплению дефектного белкового продукта. Продукт гена *Slc25a36l1* содержит домен, сходный с суперсемейством митохондриальных переносчиков, которые участвуют в транспорте пиримидиновых нуклеотидов для синтеза митохондриальной ДНК. Таким образом, SNP в гене *Slc25a36l1* может быть вовлечен в дисфункцию митохондрий, которая рассматривается как вероятная причина преждевременного старения крыс OXYS.

В ожидаемом белковом продукте гена *Rims2* у крыс OXYS отсутствуют несколько сайтов фосфорилирования серина. Этот ген ассоциирован с наследственными заболеваниями сетчатки и риском возраст-зависимых заболеваний. Мутации в гене *Lemd2*, кодирующем белок ядерной оболочки, могут приводить к гибели эмбриона или к формированию фенотипа, который включает ювенильную катаракту, наследственную дилатационную кардиомиопатию, ремоделирование хроматина и ускоренное старение сердечной ткани. Недавно показано, что мутации в гене LEMD2 могут приводить к развитию прогероидного фенотипа у людей. Полиморфизм в этом гене у крыс OXYS потенциально приводит к сдвигу рамки считывания, преждевременному стоп-кодону и потенциальной потере одного из двух трансмембранных доменов. Как показал вестерн-блот анализ, в префронтальной коре и гиппокампе крыс OXYS присутствует только полноразмерный белковый продукт гена *Lemd2*, хотя его уровень снижен по сравнению с крысами Wistar на протяжении всей жизни.

В соответствие с алгоритмом SIFT, 33 SNPs крыс OXYS, отсутствующие у нормотензивных линий, предположительно способны приводить к изменению структуры и/или функции белка и могут рассматриваться как вероятные кандидаты, с которыми связано развитие фенотипа преждевременного старения. Среди них два гена (*Acat2* и *Ephx1*) ассоциированы со старением, группа из 10 генов (*Plagr1*, *Zmym6*, *Trappc9*, *Nqo2*, *Ephx1*, *Ano10*, *Chrna5*, *Man2c1*, *Pcm1* и *LOC100364500*) связана с нейродегенеративными заболеваниями, и ген *Ephx1* – с артериальной гипертензией.

Анализ генетического сходства, основанный на идентифицированных SNP, показал, что генетически крысы OXYS наиболее близки к линии крыс ISIAH/Icgn (она также происходит из крыс Wistar/Icgn), чем к другим гипертензивным линиям

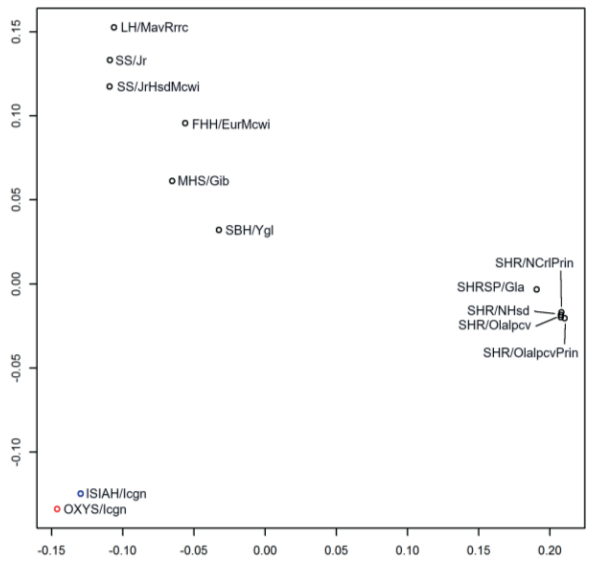


Рис. 1. Анализ расстояний между крысами OXYS/Icgn и 12 гипертензивными линиями крыс

и сублиниям крыс (рис. 1). Среди 2105 SNPs, выявленных у крыс OXYS, 725 встречались также у одной или нескольких гипертензивных линий крыс. Из них 663 были общими для крыс OXYS и НИСАГ и не были обнаружены у других крыс с гипертонией. Среди генов с данными мутациями наиболее интересен ген эпоксидгидролазы 1, который ассоциирован с гипертонией, а также коррелирует с нейродегенеративными заболеваниями и психическими расстройствами. SNP в генах *Pla2r1* и *Ccdc28b* являются наиболее распространенными у гипертензивных крыс и расположены в генетических локусах, ассоциированных с гипертонией. Высокий титр PLA2R1 в сыворотке обнаружен у пациентов с диабетом 1 типа, диабетической ретинопатией, артериальной гипертензией и нефротическим синдромом. Белок, кодируемый геном *Ccdc28b*, ассоциирован с синдромом Барде–Бидля, аутосомно-рецессивным расстройством, включающим дегенерацию сетчатки, ожирение, когнитивные нарушения, умственную отсталость, почечные расстройства и гипертензию.

Выводы: Таким образом, выявленные у крыс OXYS SNPs могут способствовать как развитию у крыс OXYS умеренной гипертензии, так и фенотипа преждевременного старения в целом, в том числе – признаков БА и ВМД. В то же время мы не выявили мутаций, которые могли бы однозначно определять данный фенотип, что свидетельствует либо о совокупном действии множественных генных нарушений, либо о факторах, не связанных с мутациями в кодирующих регионах генома крыс OXYS. Линия крыс OXYS и другие исследованные нами линии имеют мало общих мутаций, что указывает на собственные механизмы развития патологии у крыс OXYS, отличающие их даже от генетически близких крыс ISIAH/Icgn.

Финансирование: Исследование поддержано государственным бюджетным проектом (№ FWNR-2022-0016).

Single nucleotide polymorphisms associated with the development of accelerated aging phenotype in OXYS rats

Devyatkin V.*, Redina O., Kolosova N., Muraleva N.

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

** devyatkin@bionet.nsc.ru*

Key words: OXYS rats; SNP; hypertension; Alzheimer's disease; aging; age-related diseases

Motivation and Aim: Aging is a major risk factor for age-related diseases, the prevalence of which is increasing as life expectancy increases. Their development at an earlier age is a manifestation of premature aging; later development becomes the basis for successful aging and longevity. More than 60 % of people over the age of 65 suffer from several diseases, the "set" and age of manifestation of which are determined by genetic and environmental factors and quality of life. It is known that arterial hypertension, which affects 7 % of people under the age of 40 and two thirds of people over 60, contributes to the development of Alzheimer's disease (AD) and age-related macular degeneration (AMD) – incurable neurodegenerative diseases that often develop simultaneously. The genetic basis of predisposition to the development of comorbid age-related conditions remains unclear. A productive approach to studying its nature is the use of biological models, the genetic homogeneity of which and standard maintenance conditions reduce the influence of external factors and increase the reproducibility of results. The goal of this work was to search for nucleotide variants that are associated with the early development of a phenotype similar to human geriatric diseases in OXYS rats (ICG SB RAS), a unique model of accelerated aging. With mildly elevated blood pressure, OXYS rats develop cataracts, cardiomyopathy, retinopathy similar to AMD in humans, and key features of AD in the absence of mutations in the *App*, *Psen1*, and *Psen2* genes specific to the inherited form of the disease.
Methods and Algorithms: We analyzed RNA-seq data obtained previously from samples of the prefrontal cortex, hippocampus, and retina of OXYS rats. Reads were mapped to the Rnor 6.0 reference genome assembly. From the identified polymorphic variants with a minimum mapping quality of 100, positions were selected whose genotype was in a homozygous state in at least three animals in one experimental group. Reading depth (DP) must be ≥10 in at least one animal.
The SNPs identified in OXYS rats were compared with genomic and transcriptomic sequencing data from 45 other rat strains; comparisons were made only for loci sequenced in the transcriptomic analysis of OXYS rats.
Results: Comparison of transcriptome data from the prefrontal cortex, hippocampus and retina of OXYS rats with the reference genome of BN/NHsdMcwi rats revealed 42478 SNPs in 9903 genes, of which 40373 SNPs in 9699 genes were also identified in the genome sequences of 11 strains/substrains used as normotensive controls, and 22 strains/substrains of rats with a phenotype not associated with arterial hypertension or aging. Among the remaining 2105 SNPs, a search was conducted for genetic variants that may influence the manifestation of the accelerated aging phenotype in OXYS rats. Genes carrying mutations in OXYS rats are associated with antigen processing and presentation, cell growth during development, positive regulation of protein localization in the cell, regulation of binding proteins, processes in a multicellular organism, cellular senescence, major histocompatibility complex I and phagosomes.
Seven nucleotide substitutions were identified that can have a significant effect on the structure of the transcript. The most significantly truncated proteins are expected for *Rims2*, *Lemd2*, *Slc25a36l1* and *AABR07045405.1*; in the first three of them, loss of functionally significant domains is expected. Frameshift occurs in the first exon of the last two genes, which can also lead to the accumulation of a defective protein product. The *Slc25a36l1* gene product contains a domain similar to the superfamily of mitochondrial transporters that are involved in the transport of pyrimidine nucleotides for mitochondrial DNA synthesis. Thus, the SNP in the *Slc25a36l1* gene may be

involved in mitochondrial dysfunction, which is considered a likely cause of accelerated aging in OXYS rats.

The expected protein product of the *Rims2* gene in OXYS rats lacks several serine phosphorylation sites. This gene is associated with hereditary retinal diseases and the risk of age-related diseases. Mutations in the *Lemd2* gene, which encodes a nuclear envelope protein, can lead to embryonic death or a phenotype that includes juvenile cataracts, hereditary dilated cardiomyopathy, chromatin remodeling and accelerated aging of cardiac tissue. Recently, it was shown that mutations in the LEMD2 gene can lead to the development of a progeroid phenotype in humans. Polymorphisms in this gene in OXYS rats potentially result in a frameshift, a premature stop codon, and the potential loss of one of the two transmembrane domains. Western blot analysis showed that only the full-length protein product of the *Lemd2* gene is present in the prefrontal cortex and hippocampus of OXYS rats, although its level is reduced compared to Wistar rats throughout life.

According to the SIFT algorithm, 33 SNPs of OXYS rats, absent in normotensive strains, are presumably capable of leading to changes in protein structure and/or function and can be considered as likely candidates associated with the development of the accelerated aging phenotype. Among them, two genes (*Acat2* and *Ephx1*) are associated with aging, a group of 10 genes (*Plagr1*, *Zmym6*, *Trappc9*, *Nqo2*, *Ephx1*, *Ano10*, *Chrna5*, *Man2c1*, *Pcm1* and *LOC100364500*) are associated with neurodegenerative diseases, and the *Ephx1* gene is associated with arterial hypertension.

Genetic similarity analysis based on identified SNPs showed that OXYS rats are genetically closest to the ISIAH rat strain (also derived from Wistar/Icgn rats) than to other hypertensive rat strains and substrains (Fig. 1). Among the 2105 SNPs identified in OXYS rats, 725 were also found in one or more hypertensive rat strains. Of these, 663 were common to OXYS and ISIAH rats and were not found in other hypertensive rats.

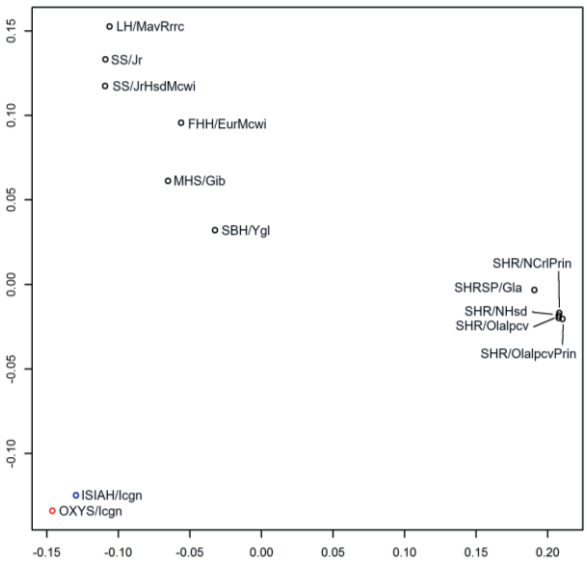


Fig. 1. Analysis of distances between OXYS/Icgn rats and 12 hypertensive rat strains

Among the genes with these mutations, the most interesting is the epoxide hydrolase 1 gene, which is associated with hypertension and also correlates with neuro-degenerative diseases and mental disorders. SNPs in the *Pla2r1* and *Ccdc28b* genes are the most common in hypertensive rats and are located in genetic loci associated with hypertension. High serum titers of PLA2R1 were found in patients with type 1 diabetes, diabetic retinopathy, hypertension, and nephrotic syndrome. The protein encoded by the *Ccdc28b* gene is associated with Bardet–Biedl syndrome, an autosomal recessive disorder involving retinal degeneration, obesity, cognitive impairment, mental retardation, renal disorders and hypertension.

Conclusion: Thus, the SNPs identified in OXYS rats may contribute to both the development of moderate hypertension in OXYS rats and the accelerated aging phenotype in general, including signs of AD and AMD. At the same time, we did not identify mutations that could unambiguously determine this phenotype, which indicates either the cumulative effect of multiple gene disorders or factors not associated with mutations in the coding regions of the OXYS rat genome. The OXYS rat strain and other strains we studied have few common mutations, which indicates their own mechanisms for the development of pathology in OXYS rats, distinguishing them even from genetically similar ISIAH/Icgn rats.
Funding: This study was supported by the state budget project (No. FWNR-2022-0016).

doi 10.18699/bgrs2024-13-09

Голый землекоп демонстрирует повышенную эффективность репарации ДНК и устойчивость к генотоксическим воздействиям

Евдокимов А.[1], Кутузов М.[1], Петрусева И.[1]*, Попов А.[1], Романенко С.[2], Перельман П.[2], Прокопов Д.[2], Селуанов А.[4], Горбунова В.[4], Графодатский А.[2], Трифонов В.[2], Ходырева С.[1], Рябчикова Е.[1], Коваль О.[1], Лаврик И.[3], Лаврик О.[1]

[1] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

[2] *Институт молекулярной и клеточной биологии СО РАН, Новосибирск, Россия*

[3] *Translational Inflammation Research, Medical Faculty, Center of Dynamic Systems, Otto von Guericke University Magdeburg, Magdeburg 39106, Germany*

[4] *Department of Biology, University of Rochester, Rochester, NY 14627, USA*

* *irapetru@niboch.nsc.ru*

Ключевые слова: *Heterochephalus glaber*; репарация ДНК; поли(АДФ-рибоза)полимераза 1; регулируемая гибель клеток

Мотивация и цель: Голый землекоп, *Heterocephalus glaber*, является самым долгоживущим видом грызунов и чрезвычайно устойчив к раку и болезням, связанным со старением. Это животное отличается рекордно высокой для грызунов максимальной продолжительностью жизни, повышенной устойчивостью к раковым заболеваниям и рядом других уникальных фенотипических черт; его геном и протеом отличаются стабильностью и эффективностью функционирования. Все это давало основание предполагать, что молекулярные «машины», противостоящие накоплению повреждений в геноме H. glaber, включая механизмы репарации ДНК, функционируют с высокой эффективностью. Механизмы регулируемой гибели клеток могут быть напрямую связаны с устойчивостью к возникновению индуцированных раковых опухолей. Однако вопрос о том, насколько эффективно и какие механизмы регулируемой клеточной гибели и как реагируют на генотоксические воздействия клетки *H. glaber* остается предметом изучения.

Методы и алгоритмы: Культивирование клеток и условия УФ-облучения описаны в [1]. Активность NER *in vitro*, эффективность удаления участка модельной ДНК-субстрата, содержащей объемное повреждение, определяли с использованием 3'-концевого мечения [1] или используя метод количественной ПЦР [2].

Результаты: Как базальная, так и УФ-индуцированная эксцизионная активность NER оказались выше в *H. glaber*, чем в клетках мыши. Более эффективное удаление урацила наблюдалось в экстрактах клеток *H. glaber* по сравнению с мышиными. Базовые уровни расщепления АР-сайта были одинаковыми в экстрактах необработанных клеток *H. glaber* и клеток мыши, но клетки ЯМР продемонстрировали более сильный ответ на обработку УФ-светом, чем клетки мыши. Для экстракта клеток *H. glaber* в моменты времени после облучения наблюдалось четко выраженное увеличение активности элонгации ДНК-полимеразы по сравнению с необлученными клетками. Уровень синтеза ПАР был

выше в экстрактах *H. glaber*, чем в клетках мыши. Изучение влияние метилметансульфоната, 5-фторурацила и этопозида на оба типа клеток с помощью современных анализов гибели клеток включая проточную цитометрию и электронную микроскопию показало, что фибробласты кожи *H. glaber* более устойчивы к действию, повреждающих ДНК агентов, чем фибробласты. *Mus musculus*. Можно также заключить, что клетки H. *glaber* демонстрируют слабо выраженный апоптотический ответ и предположительно подвергаются регулируемому некрозу. В совокупности это исследование дает новое понимание механизмов гибели клеток *H.glaber* и расширяет наше понимание основ долголетия.

Заключение. В совокупности это исследование впервые на примере путей эксционной репарации нуклеотидов и оснований демонстрирует, что удаление повреждений ДНК в клетках долгоживущего *Heterocephalus glaber* протекает более эффективно, чем в клетках коротко живущей *Mus musculus*, а также помогает продвинуться в понимании механизмов гибели клеток *H. glaber*.

Финансирование: Работа выполнена в рамках государственного задания ИХБФМ СО РАН № 121031300041-4.

Naked mole rat demonstrates increased DNA repair efficiency and resistance to genotoxic effects

Evdokimov A.[1], Kutuzov M.[1], Petruseva I.[1]*, Popov A.[1], Romanenko S.[2], Perelman P.[2], Prokopov D.[2], Seluanov A.[4], Gorbunova V.[4], Graphodatsky A.[2], Trifonov V.[2], Khodyreva S.[1], Ryabchikova E.[1], Koval O.[1], Lavrik I.[3], Lavrik O.[1]

[1] *Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia*

[2] *Institute of Molecular and Cellular Biology, Novosibirsk, Russia*

[3] *Translational Inflammation Research, Medical Faculty, Center of Dynamic Systems, Otto von Guericke University Magdeburg, Magdeburg, Germany*

[4] *Department of Biology, University of Rochester, Rochester, NY, USA*

* *irapetru@niboch.nsc.ru*

Key words: *Heterochephalus glaber*; DNA repair; poly(ADP-ribose)polymerase 1; regulated cell death

Motivation and Aim: The naked mole rat, *Heterocephalus glaber*, is the longest living rodent species and is extremely resistant to cancer and diseases associated with aging. This animal is distinguished by a record-breaking maximum life expectancy for rodents, increased resistance to cancer, and a number of other unique phenotypic traits; its genome and proteome are characterized by stability and efficient functioning. All this gave reason to assume that the molecular "machines" that resist the accumulation of damage in the *H. glaber* genome, including DNA repair mechanisms, function with high efficiency. Mechanisms of regulated cell death may be directly related to resistance to induced cancer tumors. However, the question of how effective and what mechanisms of regulated cell death and how *H. glaber* cells respond to genotoxic effects remains a subject of study.

Methods and Algorithms: The cells cultivation and UVC-light irradiation conditions described in [1]. The efficiency of removal of model DNA-substrate region containing lesions was determined using the 3′-end-labeling [1] or using qPCR-assay.

The assays of uracil excision activity of cell extracts, AP-site cleavage activity of cell extracts and PARylation with cell extract proteins described in [2].
Metabolic assay, cell death detection and caspase activity detection described in [2].
Results: Both basal and UV-induced NER excision activities were higher in *H. glaber* than in mouse cells. More efficient removal of uracil was observed in *H. glaber* cell extracts compared to mouse extracts. Baseline levels of AP site cleavage were similar in extracts of untreated *H. glaber* and mouse cells, but NMR cells showed a stronger response to UV light treatment than mouse cells. For the *H. glaber* cell extract, a clear increase in DNA polymerase elongation activity was observed at time points after irradiation compared to non-irradiated cells. The level of PAR synthesis was higher in *H. glaber* extracts than in mouse cells. Studying the effects of methylmethanesulfonate, 5-fluorouracil and etoposide on both cell types using modern cell death assays, including flow cytometry and electron microscopy, showed that *H. glaber* skin fibroblasts are more resistant to DNA-damaging agents than *Mus musculus* fibroblasts. It can also be concluded that *H. glaber* cells exhibit a mild apoptotic response and presumably undergo regulated necrosis. Taken together, this study provides new insight into the mechanisms of *H. glaber* cell death and expands our understanding of the basis of longevity.
Conclusion: Taken together, this study demonstrates for the first time, using the example of nucleotide and base excision repair pathways, that the removal of DNA damage in the cells of the long-lived *Heterocephalus glaber* is more efficient than in the cells of the short-lived *Mus musculus*, and also helps to advance the understanding of the mechanisms of cell death *H. glaber*.
Funding: The study is supported by the Ministry of Science and Higher Education of Russian Federation (project number 121031300041-4).

Список литературы/References

1. Evdokimov A., Kutuzov M., Petruseva I. et al. Naked mole rat cells display more efficient excision repair than mouse cells. *Aging (Albany NY)*. 2018;10(6):1454-1473. doi 10.18632/aging.101482
2. Evdokimov A., Popov A., Ryabchikova E. et al. Uncovering molecular mechanisms of regulated cell death in the naked mole rat. *Aging (Albany NY)*. 2021;13(3):3239-3253. doi 10.18632/aging.202577

doi 10.18699/bgrs2024-13-10

Характеристика параметров продолжительности жизни и стрессоустойчивости у *Drosophila melanogaster* с повсеместной и нейрональной сверхэкспрессией генов репарации ДНК

Коваль Л.А.[1]*, Прошкина Е.Н.[1], Москалев А.А.[1, 2]

[1] *Институт биологии Коми научного центра УрО РАН, Сыктывкар, Россия*

[2] *Институт биологии старения, Нижний Новгород, Россия*

* *koval.l.a@ib.komisc.ru*

Ключевые слова: старение; *Drosophila*; гены репарации ДНК; продолжительность жизни; стрессоустойчивость

Мотивация и цель: С возрастом происходит нарушение эффективности репарации ДНК [1]. Одной из причин этого является снижение экспрессии ферментов репарации ДНК или их активности. Мы предположили, что сверхэкспрессия генов репарации ДНК приведет к повышению длительности жизни и устойчивости дрозофил к различным видам стресса. Для этого исследовали роль повсеместной и нейрональной сверхэкспрессии генов репарации ДНК на продолжительность жизни (ПЖ) и устойчивость *Drosophila melanogaster* к различным повреждающим факторам окружающей среды (голодание, гипертермия, окислительный стресс). Сверхэкспрессировали гены, которые участвуют в распознавании повреждений ДНК (гомологи *HUS1, CHK2, GADD45*), репарации эксцизии нуклеотидов и оснований (гомологи *XPF, XPC* и *APE1*), а также репарации двуцепочечных разрывов ДНК (гомологи *BRCA2, XRCC3, KU80* и *WRNexo*).

Методы и алгоритмы: Культивировали и содержали дрозофил в стандартных условиях, описанных в [2]. Использовали линии с дополнительными копиями генов репарации ДНК под контролем промотера UAS: *w^{1118},UAS-D-Gadd45; w^{1118},UAS-Brca2; w^{1118},UAS-Hus1; w^{1118},UAS-mnk; w^{1118},UAS-spn-B,* любезно предоставленные др. Ури Абду (Университет Бен-Гуриона в Негеве, Беер-Шева, Израиль), *w^{1118},UAS-mei-9; w^{1118},UAS-mus210; w^{1118},UAS-Ku80; w^{1118},UAS-Rrp1; w^{1118},UAS-WRNexo*, были заказаны у компании GenetiVision (GenetiVision Houston, США). Линии с кондиционным (миферистон-индуцибельным) драйвером GAL4-GeneSwitch: *P{Act5C(-FRT) GAL4.Switch.PR}3/TM6B, Tb1* повсеместный и *y- w*; P{ELAV-GeneSwitch}* нейрональный (BDSC, Университет Индианы, США). Для активации сверхэкспрессии использовали RU486-индуцибельную систему GeneSwitch [3], и систему GAL4/UAS, которая использует дрожжевой активатор транскрипции GAL4 и вышестоящую активирующую последовательность UAS, с которой он связывается, для управления экспрессией трансгена [4]. Анализ стрессоустойчивости и ПЖ вели по ранее описанным протоколам [2, 5].

Результаты: Изучили влияние серхэкспрессии генов репарации ДНК на параметры ПЖ самцов и самок *Drosophila melanogaster*. Обнаружили увеличение ПЖ при нейрональной сверхактивации генов *Hus1* у самцов и самок и *mei-9* (гомолог *XPF*) у самцов. В то же время повсеместная сверхэкспрессия значительно снизила длительность жизни и параметры ПЖ. Как правило, ПЖ сопряжена с увеличением устойчивости к различным видам стресса. Поскольку гены репарации ДНК вовлечены в механизмы ответа на действие экзогенных и эндогенных факторов, мы предположили, что их сверхэкспрессия приведет к

повышению устойчивости дрозофил к различным видам стресса. Анализ стрессоустойчивости выявил улучшение выживаемости при гипертермии у дрозофил с нейрональной сверхэкспрессия генов *Rrp1* (гомолог *APE1*) у самцов, *Hus1* и *mei-9* у самок. В условиях окислительного стресса выживаемость повышалась у самок со сверхавтивацией гена *mei-9*. Наиболее устойчивы к голоданию оказались самцы с нейрональной сверхэкспрессией гена *Rrp1* и самки со сверхактивацией гена *Ku80*. Повсеместная сверхэкспрессия также повышала устойчивость к гипертермии у трансгенных дрозофил с индукцией генов *Brca2, Hus1, Rrp1* у самцов и *mnk* (гомолог *chk2*) у самок. В условиях окислительного стресса были более устойчивы самцы со сверхэкспрессией генов *mus210* (гомолог *XPC*) и *spn-B* (гомолог *XRCC3*), у самок – *D-Gadd45, Hus1, spn-B*. При голодании самцы с повсеместной сверхактивацией генов *Brca2, D-Gadd45, spn-B и WRNexo*.
Выводы: Мы исследовали параметры продолжительности жизни и стрессоустойчивости у трансгенных особей *Drosophila melanogaster* со сверхэкспрессией генов, которые участвуют в репарации ДНК. Обнаружили как положительные, так и отрицательные эффекты в зависимости от пола, ткани, в которой происходила индукция сверхэкспрессии и роли гена в процессе репарации ДНК.
Финансирование: Исследование выполнено в рамках государственного задания по теме «Генетические и функциональные исследования эффектов геропротекторных вмешательств на модели *Drosophila melanogaster*» № 122040600022-1.

Characterization of longevity and stress resistance parameters in *Drosophila melanogaster* with ubiquitous and neuronal overexpression of DNA repair genes

Koval L.A.[1]*, Proshkina E.N.[1], Moskalev A.A.[1, 2]

[1] *Institute of Biology of Komi Science Center of Ural Branch of Russian Academy of Sciences, Syktyvkar, Russia*

[2] *Institute of Biology of Ageing, Nizhny Novgorod, Russia*

* *koval.l.a@ib.komisc.ru*

Key words: aging; Drosophila; DNA repair gene; lifespan; stress resistance

Motivation and Aim: DNA repair declines with age [1]. One of the reasons for this is a decrease in the expression of DNA repair enzymes or their activity. We hypothesized that overexpression of DNA repair genes would lead to increased longevity and resistance of *Drosophila melanogaster* to various types of stress. Here we investigated the role of ubiquitous and neuronal overexpression of DNA repair genes on lifespan (LS) and resistance of Drosophila melanogaster to various damaging environmental factors (starvation, hyperthermia, oxidative stress). Overexpressed genes included those involved in recognition of DNA damage (homologs of HUS1, CHK2, GADD45), nucleotide and base excision repair (homologs of XPF, XPC and AP-endonuclease-1), and repair of double-stranded DNA breaks (homologs of BRCA2, XRCC3, KU80 and WRNexo).
Methods and Algorithms: Drosophila were cultured and maintained under standard conditions described in [2]. We used strains with additional copies of DNA repair genes under UAS control**:** *w^{1118},UAS-Brca2; w^{1118},UAS-Hus1; w^{1118},UAS-mnk; w^{1118},UAS-spn-*

B; w^{1118},UAS-D-Gadd45, kindly provided by Dr. Uri Abdu (Ben-Gurion University of the Negev, Beer-Sheva, Israel). *w^{1118},UAS-mei-9; w^{1118},UAS-mus210; w^{1118},UAS-Ku80; w^{1118},UAS-Rrp1; w^{1118},UAS-WRNexo*, were ordered from GenetiVision (GenetiVision Houston, USA). Strains with a conditioned (miferistone-inducible) GAL4-GeneSwitch driver: *P{Act5C(-FRT) GAL4.Switch.PR}3/TM6B, Tb1* ubiquitous and *y- w*; P{ELAV-GeneSwitch}* neuronal (BDSC, Indiana University, USA). The RU486-inducible GeneSwitch system [3] was used to activate overexpression, and the GAL4/UAS system, which uses the yeast activator of transcription GAL4 and the upstream activating sequence of UAS to which it binds to control transgene expression [4]. Stress resistance and LS were analyzed according to previously described protocols [2, 5].

Results: The effect of DNA repair gene overexpression on the parameters of Drosophila melanogaster's were studied. We found an increase in LS parameters during neuronal overactivation of *Hus1* genes in males and females and *mei-9* (*XPF* homologue) in males. At the same time, ubiquitous overexpression significantly reduced LS parameters. Generally, LS is associated with increased resistance to various types of stress. Since DNA repair genes are involved in response mechanisms to exogenous and endogenous factors, we hypothesized that their overexpression would result in increased resistance of Drosophila to stress. Analysis of stress resistance revealed improved survival under hyperthermia in *Drosophila* with neuronal overexpression of the genes *Rrp1* (*APE1* homologue) in males, *Hus1* and *mei-9* in females. Under conditions of oxidative stress, survival was increased in females with *mei-9* gene overactivation. Males with neuronal overexpression of the *Rrp1* gene and females with overactivation of the *Ku80* gene were the most resistant to starvation. Ubiquitous overexpression also increased resistance to hyperthermia in transgenic *Drosophila* with induction of *Brca2*, *Hus1*, and *Rrp1* genes in males and *mnk* (*chk2* homologue) in females. Under oxidative stress, males with overexpression of *mus210* (*XPC* homologue) and *spn-B* (*XRCC3* homologue) genes were more resistant, while *D-Gadd45*, *Hus1*, and *spn-B* genes were more resistant in females. Under starvation, males with ubiquitous overactivation of *Brca2, D-Gadd45, spn-B*, and *WRNexo* genes.

Conclusion: We investigated the potential life-extending effects of increasing the expression of genes known to be involved in DNA repair in *Drosophila*. We found both positive and negative effects depending on sex, tissue in which overexpression was induced and the role of the gene in the DNA repair process.

Funding: The study was carried out within the framework of the state task on the theme «Genetic and functional studies of the effects of geroprotective interventions on the *Drosophila melanogaster* model» No. 122040600022-1.

Список литературы/References

1. Gorbunova V. et al. Changes in DNA repair during aging. *Nucleic Acids Res*. 2007;35(22):7466-7474. doi 10.1093/nar/gkm756
2. Koval L. et al. The role of DNA repair genes in radiation-induced adaptive response in Drosophila melanogaster is differential and conditional. *Biogerontology*. 2020;21(1):45-56. doi 10.1007/s10522-019-09842-1
3. Osterwalder T. et al. A conditional tissue-specific transgene expression system using inducible GAL4. *Proc Natl Acad Sci USA*. 2001;98(22):12596-12601. doi 10.1073/pnas.221303298
4. Weitzman J.B. Switching on genes with GeneSwitch. *Genome Biol*. 2001;2:spotlight-20011029-01. doi 10.1186/gb-spotlight-20011029-01
5. Moskalev A. et al. The role of D-GADD45 in oxidative, thermal and genotoxic stress resistance. *Cell Cycle*. 2012;11(22):4222-4241. doi 10.4161/cc.22545

doi 10.18699/bgrs2024-13-11

Исследование влияния производного 7-гидроксикумарина UB3-C_{10} на хронологическое старение миобластов человека MB135 *in vitro*

Моргунова Г.В.[1]*, Челомбитько М.А.[2], Лямзаев К.Г.[2]

[1] *Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

[2] *ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России, Российский геронтологический научно-клинический центр, Москва, Россия*

* *morgunova@mail.bio.msu.ru*

Ключевые слова: геропротекторы; мышечные клетки; биомаркеры старения; митофагия

Цель: Поиск потенциальных геропротекторных препаратов, в том числе на клеточных культурах является актуальной задачей в геронтологии. Важную роль в развитии возрастных патологий играют дисфункции митохондрий [1]. Активация митофагии является одним из способов удаления дисфункциональных митохондрий [2], он особенно важен для долгоживущих клеток — таких как нейроны и миоциты [3]. Активацию митофагии можно вызывать разобщением окислительного фосфорилирования и дыхания. Производное 7-гидроксикумарина UB3-C_{10} является относительно сильным разобщителем и наряду с этим показывает низкую токсичность, в том числе в отношении нейронов [4]. Целью настоящей работы стало исследование геропротекторных свойств UB3-C_{10} в модели хронологического старения мышечных клеток человека.

Методы: Исследование проводили на нормальных иммортализованных миобластах MB135, полученных из биопсии мышц здорового человека. Клетки MB135 культивировали в течение 1 и 4 недель с добавлением UB3-C_{10} в концентрациях 2, 10 и 50 мкМ. Для определения сенесцентного состояния клеток были использованы следующие методы: гистохимическое выявление ассоциированной со старением β-галактозидазы (SA-β-гал), иммуноферментная оценка уровня провоспалительных цитокинов ИЛ-6 и ИЛ-8, а также определение белков p21 и pRB (Ser807/811) с помощью вестерн-блоттинга. Помимо этого, изучали влияние UB3-C_{10} на количество митохондрий в клетках, для этого определяли содержание митохондриальных белков COX-IV и TOM20 с помощью вестерн-блоттинга через 24, 48 и 72 ч инкубации клеток с UB3-C_{10} в концентрациях 2, 10 и 50 мкМ.

Результаты: Ранее нами было показано, что культивирование клеток MB135 в течение 4 недель можно рассматривать как одну из моделей для изучения старения *in vitro* [4]. Добавление UB3-C_{10} в культуральную среду в концентрации 50 мкМ приводило к снижению доли SA-β-гал-положительных клеток и уровня белка p21, а также увеличению уровня фосфорилированного pRb (Ser807/811) в клетках. Полученные данные свидетельствуют о геропротекторных свойствах UB3-C_{10}, в то же время не было обнаружено влияние UB3-C_{10} на секрецию провоспалительных цитокинов. Поскольку известно, что накопление дисфункциональных митохондрий усугубляет клеточное старение, в основе геропротекторного эффекта UB3-C_{10} может лежать обновление пула митохондрий

под его воздействием. В связи с этим мы оценили способность UB3-C_{10} индуцировать митофагию в мышечных клетках. Для этого мы определяли содержание митохондриальных белков COX-IV и TOM20. Инкубация клеток с UB3-C_{10} в течение 24 ч приводит к дозозависимому снижению количества митохондрий, а спустя 72 ч – количество митохондрий возвращается к уровню контрольных значений. С помощью флуоресцентной микроскопии также была детектирована митофагия в клетках, инкубированных UB3-C_{10} в концентрации 50 мкМ в течение 24 ч. Полученные данные указывают на то, что UB3-C_{10} действительно может вызвать митофагию, однако механизмы этого эффекта требуют дальнейшего изучения.
Выводы: На основании полученных данных можно предположить, что производное 7-гидроксикумарина UB3-C_{10} проявляет геропротекторные свойств в модели хронологического старения мышечных клеток человека линии MB135, при этом защитный эффект может быть связан с его влиянием на пул митохондрий.
Финансирование: Исследование поддержано РНФ (№ 23-14-00061).

Investigation of the effect of the 7-hydroxycoumarin derivative UB3-C_{10} on chronological aging of human myoblasts MB135 *in vitro*

Morgunova G.V.[1]*, Chelombitko M.A.[2], Lyamzaev K.G.[2]
[1] *Lomonosov Moscow State University, Moscow, Russia*
[2] *Russian Clinical and Research Center for Gerontology, Ministry of Healthcare of the Russian Federation, Pirogov Russian National Research Medical University, Moscow, Russia*
* *morgunova@mail.bio.msu.ru*

Key words: geroprotectors; muscle cells; biomarkers of aging; mitophagy

Aim: The search for potential geroprotectors, including on cell cultures, is an urgent area of gerontological researches. Mitochondrial dysfunction plays an important role in the development of age-related pathologies [1]. Mitophagy activation is one of the ways to remove dysfunctional mitochondria [2], it is crucial for long-lived cells, such as neurons and myocytes [3]. Activation of mitophagy can be induced by uncoupling oxidative phosphorylation and respiration. The 7-hydroxycoumarin derivative UB3-C_{10} is a relatively strong uncoupler and along with this it shows low toxicity including against neurons [4]. The aim of the present work was to study the geroprotective properties of UB3-C_{10} in the model of chronological aging of human muscle cells.
Methods: The study was performed on normal immortalized MB135 myoblasts obtained from muscle biopsies of healthy humans. MB135 cells were cultured for 1 and 4 weeks with the addition of UB3-C_{10} at concentrations of 2, 10 and 50 μM. The following methods were used to determine the senescent state of the cells: histochemical detection of senescence-associated β-galactosidase (SA-β-gal), immunoassessment of pro-inflammatory cytokines IL-6 and IL-8, and determination of p21 and pRB (Ser807/811) proteins by Western blotting. In addition, we studied the effect of UB3-C_{10} on the number of mitochondria in cells, for this purpose we determined the content of

mitochondrial proteins COX-IV and TOM20 by Western blotting after 24, 48 and 72 h of incubation of cells with UB3-C_{10} at concentrations of 2, 10 and 50 μM.

Results: We have previously shown that culturing MB135 cells for 4 weeks can be considered as one of the models for studying aging *in vitro* [4]. The addition of UB3-C_{10} to the culture medium at a concentration of 50 μM resulted in a decrease in the proportion of SA-β-gal-positive cells and the level of p21 protein, as well as an increase in the level of phosphorylated pRb (Ser807/811) in cells. These findings suggest geroprotective properties of UB3-C_{10}, while no effect of UB3-C_{10} on the secretion of pro-inflammatory cytokines was detected. Since the accumulation of dysfunctional mitochondria is known to exacerbate cellular senescence, the geroprotective effect of UB3-C_{10} may be based on the renewal of the mitochondrial pool under its influence. In this regard, we evaluated the ability of UB3-C_{10} to induce mitophagy in muscle cells. For this purpose, we determined the content of mitochondrial proteins COX-IV and TOM20. Incubation of cells with UB3-C_{10} for 24 h leads to a dose-dependent decrease in the number of mitochondria, and after 72 h the number of mitochondria returns to the level of control values. Fluorescence microscopy was also used to detect mitophagy in cells incubated with UB3-C_{10} at a concentration of 50 μM for 24 hr. The data obtained indicate that UB3-C_{10} can indeed induce mitophagy, but the mechanisms of this effect require further investigation.

Conclusion: Based on the obtained data, it can be assumed that the 7-hydroxycoumarin derivative UB3-C_{10} exhibits geroprotective properties in the model of chronological aging of human muscle cell line MB135, and the protective effect may be related to its influence on the mitochondrial pool.

Funding: The study is supported by Russian Science Foundation (No. 23-14-00061).

Список литературы/References

1. Lane R.K., Hilsabeck T., Rea S.L. The role of mitochondrial dysfunction in age-related diseases. *Biochim Biophy. Acta Bioenerg*. 2015;1847(11):1387-1400
2. Ashrafi G., Schwarz T.L. The pathways of mitophagy for quality control and clearance of mitochondria. *Cell Death Differ*. 2013;20(1):31-42
3. Morgunova G.V., Shilovsky G.A., Khokhlov A.N. Effect of caloric restriction on aging: Fixing the problems of nutrient sensing in postmitotic cells? *Biochemistry (Mosc.).* 2021;86:1352-1367
4. Krasnov V.S., Kirsanov R.S., Khailova L.S., Popova L.B., Lyamzaev K.G., Firsov A.M., Korshunova G.A., Kotova E.A., Antonenko Y.N. Alkyl esters of 7-hydroxycoumarin-3-carboxylic acid as potent tissue-specific uncouplers of oxidative phosphorylation: Involvement of ATP/ADP translocase in mitochondrial uncoupling. *Arch Biochem Biophys.* 2022;728:109366
5. Chelombitko M.A., Morgunova G.V., Strochkova N.Y., Lyamzaev K.G. Geroprotective effect of the mitochondrial-targeted antioxidant SkQBerb in a model of the chronological aging of MB135 human myoblasts. *Adv Gerontol.* 2023;13(2):94-98

doi 10.18699/bgrs2024-13-12

Сравнение кинетических параметров рекомбинантных PARP1 человека и голого землекопа

Назаров К.Д.[1, 2]*, Петрусева И.О.[2], Лаврик О.И.[2]

[1] *Новосибирский государственный университет, Новосибирск, Россия*

[2] *Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия*

** k.nazarov1@g.nsu.ru*

Ключевые слова: репарация; старение; голый землекоп; PARP1

Мотивация и цель: Голый землекоп (*Heterocephalus glaber*) представляет собой перспективную модель для изучения функционирования систем поддержания стабильности генома. Этот небольшой грызун демонстрирует экстремально высокую продолжительность жизни (~32 года) по сравнению с близкой по размеру и массе тела мышью (*Mus musculus*) (~4 года). Клетки голого землекопа обладают рядом особенностей функционирования генома, транскриптома и протеома, которые обусловливают устойчивость его клеток к старению, раковым заболеваниям, а также окислительному стрессу. Кроме того, клетки голого землекопа имеют более эффективные системы репарации ДНК по сравнению с клетками близкой по размеру тела мыши [1]. Несмотря на это, особенности функционирования белков репарации, обеспечивающие повышенную эффективность репарации ДНК в клетках голого землекопа, требуют более углубленного исследования. Ключевым регулятором многих клеточных процессов, в том числе репарации ДНК, является фермент PARP1. Регуляция работы многих клеточных белков-мишеней, включая собственную активность PARP1, обеспечивается благодаря катализируемой данным ферментом реакции поли(АДФ-рибозил)ирования.

Ранее нами было проведено изучение уровня синтеза поли(АДФ-рибозы) на белковых экстрактах клеток человека (HEK 293T) и голого землекопа (NSF8) [2]. Было показано, что в экстракте клеток HEK 293T уровень аутомодификации PARP1 был в 10 раз выше по сравнению с наблюдаемым в экстракте клеток NSF8. Наблюдаемые различия могут быть обусловлены как неопределенной концентрацией PARP1 в экстрактах, так и присутствием в них белковых факторов, стимулирующих аутомодификацию PARP1. В связи с этим целью нашей работы было получение рекомбинантных PARP1 человека и голого землекопа и проведение сравнительной оценки их кинетических параметров.

Данная работа посвящена сравнительному изучению характеристик взаимодействия рекомбинантного PARP1 голого землекопа и человека с ДНК, а также его термостабильности и устойчивости к денатурирующим воздействиям другой природы. Цель работы – получение рекомбинантного PARP1 голого землекопа и сравнение ключевых характеристик (кинетические параметры реакции парилирования, сродство к ДНК, термостабильность и устойчивость к денатурирующим воздействиям), с соответствующими характеристиками PARP1 человека.

Методы и алгоритмы: Наработка PARP1 проведена в клетках *E. coli* BL21 Gen-X методом автоиндукции при 37 °C в течение 18 часов с использованием плазмиды pLate31$_{_Hgl}$PARP1; целевой белок был очищен хроматографически. Кинетические

параметры полученных рекомбинантных PARP1 определяли методом флуоресцентного титрования с измерением анизотропии флуоресценции.
Результаты: В результате данной работы была выбрана методика очистки рекомбинантного белка PARP1 голого землекопа. Было показано, что значения K_M и V_{max} рекомбинантных PARP1 человека и голого землекопа близки, что говорит о схожести кинетических свойств двух белков. Сравнительный биоинформатический анализ аминокислотной последовательности обнаружил 13 уникальных аминокислотных замен в структуре PARP1 голого землекопа по сравнению с PARP1 человека и мыши. Однако ни одна из замен не затрагивает каталитическую триаду His862, Tyr896 и Glu988.
Выводы: Результаты биоинформатического анализа соответствуют полученным экспериментальным путем значениям кинетических параметров для очищенных рекомбинантных PARP1 человека и голого землекопа, экспрессированных в *E. coli*.
Финансирование: Работа выполнена в рамках государственного задания ИХБФМ СО РАН № 121031300041-4.

Comparison of kinetic parameters of recombinant PARP1 man and naked mole rat

Nazarov K.D.[1, 2]*, Petruseva I.O.[2], Lavrik O.I.[2]

[1] *Novosibirsk State University, Novosibirsk, Russia*

[2] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*

** k.nazarov1@g.nsu.ru*

Key words: reparation; aging; naked mole rat; PARP1

Motivation and Aim: The naked mole rat (*Heterocephalus glaber*) represents a promising model for studying the functioning of genome stability maintenance systems. This small rodent demonstrates an extremely high life lifespan (~32 years) compared to a mouse (*Mus musculus*) of similar size and body weight (~4 years). Naked mole rat cells have a number of features of the functioning of the genome, transcriptome and proteome, which determine the resistance of its cells to aging, cancer, and oxidative stress. In addition, naked mole rat cells have more efficient DNA repair systems than cells of a similar body size in a mouse [1]. Despite this, the peculiarities of the functioning of repair proteins, which ensure increased efficiency of DNA repair in naked mole rat cells, require more in-depth study. A key regulator of many cellular processes, including DNA repair, is the enzyme PARP1. Regulation of the work of many cellular target proteins, including the intrinsic activity of PARP1, is ensured by the poly(ADP-ribosyl)ation reaction catalyzed by this enzyme.
Previously, we studied the level of poly(ADP-ribose) synthesis on protein extracts of human cells (HEK 293T) and naked mole rat (NSF8) [2]. It was shown that in the HEK 293T cell extract the level of PARP1 automodification was 10 times higher compared to that observed in the NSF8 cell extract. The observed differences may be due to both an uncertain concentration of PARP1 in the extracts and the presence of protein factors in them that stimulate PARP1 automodification. In this regard, the goal of this work was to obtain recombinant human and naked mole rat PARP1, and to conduct a comparative assessment of their kinetic parameters.

This work is devoted to a comparative study of the characteristics of the interaction of recombinant naked mole rat and human PARP1 with DNA, as well as its thermal stability and resistance to denaturing influences of other nature. Purpose of the work: to obtain recombinant naked mole rat PARP1 and compare key characteristics (kinetic parameters of the parylation reaction, affinity for DNA, thermal stability and resistance to denaturing influences) with the corresponding characteristics of human PARP1.
Methods and Algorithms: PARP1 production was carried out in *E. coli* BL21 Gen-X cells by autoinduction at 37 °C for 18 hours using the pLate31_HglPARP1 plasmid; the target protein was purified by chromatography. The kinetic parameters of the obtained recombinant PARP1 were determined by fluorescence titration with fluorescence anisotropy measurements.
Results: As a result of this work, a method for purifying the recombinant naked mole rat PARP1 protein was selected. It was shown that the K_M and V_{max} values of recombinant human and naked mole rat PARP1 are close, which indicates the similarity of the kinetic properties of the two proteins. Comparative bioinformatics analysis of the amino acid sequence revealed 13 unique amino acid substitutions in the structure of naked mole rat PARP1 compared to human and mouse PARP1. However, none of the substitutions affected the catalytic triad His862, Tyr896 and Glu988.
Conclusion: The results of bioinformatics analysis are consistent with experimentally obtained kinetic parameter values for purified recombinant human and naked mole rat PARP1 expressed in *E. coli*.
Funding: The work was supported by the Ministry of Science and Higher Education of Russian Federation (project number 121031300041-4).

Список литературы/References

1. Evdokimov A., Kutuzov M., Petruseva I., Lukjanchikova N. et al. Naked mole rat cells display more efficient excision repair than mouse cells. *Aging (Albany NY)*. 2018;10:1454-1473
2. Kosova A., Kutuzov M., Evdokimov A., Ilina E., Belousova E., Romanenko S., Trifonov V., Khodyreva S., Lavrik O. Poly(ADP-ribosyl)ation and DNA repair synthesis in the extracts of naked mole rat, mouse, and human cells. *Aging (Albany NY)*. 2019;11:2852-2873

doi 10.18699/bgrs2024-13-13

Исследование геропротекторных свойств уролитина А и его производного с повышенной мембранофильностью *in vitro*

Павлюченкова А.Н.[1, 2], Челомбитько М.А.[1, 2], Моргунова Г.В.[3], Строчкова Н.Ю.[4], Лямзаев К.Г.[1, 2]*

[1] *Российский клинический и научный центр геронтологии, Российский национальный исследовательский медицинский университет им. Н.И. Пирогова, Москва, Россия*

[2] *Институт физико-химической биологии им. А.Н. Белозерского, Московский государственный университет, Москва, Россия*

[3] *Биологический факультет, Московский государственный университет, Москва, Россия*

[4] *Факультет биоинженерии и биоинформатики, Московский государственный университет, Москва, Россия*

[*] *Lyamzaev@gmail.com*

Ключевые слова: уролитин А; сенесцентные клетки; геропротекторы

Цели: Митохондриальная дисфункция и клеточный сенисценс являются отличительными признаками старения и тесно взаимосвязаны. Митохондриальная дисфункция, обычно сопровождающееся повышенной продукцией свободных радикалов кислорода, является причиной и следствием клеточного старения и играет важную роль в поддержании фенотипа старения. Одним из основных механизмов утилизации нефункциональных митохондрий в клетке является митофагия и ее активаторы рассматриваются в настоящее время как перспективные средства для борьбы с рядом возраст-зависимых заболеваний [1]. Уролитин А является метаболитом эллаговой кислоты, и проявляет геропротекторные и антиоксидантные свойства в различных модельных системах [2, 3]. Известно, что Уролитин А, обладает небольшой разобщающей активность [4], но при этом способен индуцировать митофагию, способом, вероятно, не связанным с разобщением окислительного фосфорелирования [5]. Поскольку разобщение окислительного фосфорилирования, является сильным стимулятором митофагии, мы предположили, что модификация уролитина А, усиливающая его разобщающие свойства, должна усиливать его способность вызывать митофагию, а, значит, и повысить его биологическую активность. Ранее было показано, что добавление линейного алкила, содержащего 6 атомов углерода, в структуру молекулы способно усиливать ее разобщающие свойства [6]. Основываясь на этом факте, нами было синтезировано производное уролитина А с мембранофильной группой C6H13 (KIRS152). Данное соединения обладает более сильным разобщающей активность и лучше накапливается в клетках по сравнению с исходным Уролитином А. Целью нашей работы было сравнение геропротекторного эффекта Уролитина А и KIRS152 в модели цисплатин-индуцированного старения фибробластов кожи человека.

Методы: Исследование проводилось на нормальных первичных фибробластах от здоровых доноров. С помощью резазуринового теста было проведено сравнение цитотоксичности уролитина А и KIRS152 для первичных фибробластов. Далее изучали влияние нетоксичных концентраций исследуемых соединений на маркеры сенесценса, индуцированного цисплатином. Для индукции старения

фибробласты в течение 24 ч инкубировали с 10 мкМ цисплатина. После отмывки к клеткам добавляли уролитин А и KIRS152 в диапазоне концентраций от 0.5 до 100 мкМ. И инкубировали в течение 6 сут. На 7-е сут проводили анализ признаков сенесцентного фенотипа.

В качестве основного маркера старения был использован метод оценки индекса старения (ИС) с помощью проточного цитометра с системой визуализации Amnis Flowsight. ИС рассчитывали по следующей формуле: ИС = ((nAF - 1) + 5 × (nD - 1))/2, где nAF – нормализованная автофлуоресценция липофусцина, а nD – нормализованный диаметр клетки. С возрастом клетки накапливают липофусцин и увеличиваются в размере. Таким образом, чем старше клетка, тем выше ИС. Помимо этого, для оценки проявления сенесцентного фенотипа были использованы следующие методики: гистохимическая оценка проявления ассоциированной со старением β-галактозидазы (SA-β-gal) и определение содержания цитокина GDF15 (характеризует секреторный фенотип, ассоциированный со старением) с помощью иммуноферментного анализа.

Результаты: Анализ результатов резазуринового теста показал, что инкубация фибробластов с уролитином А в течение 24 часов снижает жизнеспособность фибробластов, начиная с концентрации 20 мкМ. В то время как KIRS152 нетоксичен вплоть до 100 мкМ. Анализ результатов оценки ИС на проточном цитофлуориметре показал, что как KIRS152, так и уролитин А дозозависимо снижают значение ИС в диапазоне концентраций от 2 до 25 мкМ. При этом геропротекторный эффект для KIRS152 выражен сильнее, чем для уролитина А. Поскольку считается, что геропротекторный эффект уролитина А связан с его способностью индуцировать митофагию, то подавление митофагии должно отменять этот эффект. Для того чтобы проверить это предположение, мы использовали SBI-0206965 – ингибитор киназы ULK1, играющей критическую роль в аутофагии. Предобработка фибробластов SBI-0206965 отменяло геропротекторный эффект KIRS152 и уролитина А. Что свидетельствует в пользу предположения о том, что механизм клеточного действия исследуемых веществ может быть связан с индукцией митофагии, являющейся типом селективной аутофагией. Для оценки влияния геропротекторов KIRS152 и уролитина А на проявление сенесцентного фенотипа, была измерена доля SA-β-gal-положительных клеток после гистохимического окрашивания. SA-β-gal является биомаркером активации лизосомального пути старения, для которого было показано значительное повышение экспрессии с увеличением возраста клеток. Анализ полученных данных продемонстрировал дозозависимое уменьшение доли SA-β-gal-положительных клеток как для KIRS152, так и для уролитина А в исследуемых диапазонах концентраций. Так как KIRS152 заметно снижает ИС, но оказывает меньший эффект на долю SA-β-gal-положительных клеток, можно предположить, что он мешает накоплению в клетке липофусцина. Оценка содержания в инкубационной среде фактора роста и дифференцировки GDF15, повышение уровня которого также является признаком сенесценса, показала, что KIRS152 в концентрации 2 мкМ предотвращает цисплатин-индуцированное увеличение секреции GDF15, в то время как уролитин А не оказывает подобного эффекта.

Выводы: Таким образом, мы показали, что оба соединения – KIRS152 и уролитин А – дозозависимо снижают долю SA-β-положительных клеток, а также ИС в модели цисплатин-индуцированного старения фибробластов кожи человека. В то

же время эффект KIRS152 на ИС выше, чем для уролитина А в тех же концентрациях. Предобработка фибробластов ингибитором киназы ULK1 препятствует геропротекторным свойствам обоих соединений, что свидетельствует о вкладе митофагии/аутофагии в защитный эффект уролитина А и его производного. При этом только KIRS152 предотвращает усиление секреции GDF15. Кроме того KIRS152 является менее токсичным соединением для фибробластов человека, чем уролитин А. Все вышеперечисленное указывает на то, что геропротекторные свойства KIRS152 превосходят свойства уролитина А по ряду параметров.
Финансирование: Исследование поддержано РНФ (№ 23-14-00061).

Study of geroprotective properties of urolithin A and its derivative with increased membranophilicity *in vitro*

Pavlyuchenkova A.N.[1, 2], Chelombitko M.A.[1, 2], Morgunova G.V.[3], Strochkova N.Yu.[4], Lyamzaev K.G.[1, 2]*

[1] *Russian Clinical and Research Center for Gerontology, Pirogov Russian National Research Medical University, Moscow, Russia*

[2] *Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow, Russia*

[3] *School of Biology, Moscow State University, Moscow, Russia*

[4] *Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia*

* *Lyamzaev@gmail.com*

Key words: urolithin A; senescence; geroprotectors

Aims: Mitochondrial dysfunction and cellular senescence are key features of aging and are closely intertwined. Mitochondrial dysfunction, often accompanied by heightened production of free radicals, both leads to and results from cellular senescence, playing a pivotal role in maintaining the aging phenotype. Mitophagy, the cellular process responsible for disposing of dysfunctional mitochondria, and its activators are currently being investigated as potential tools to combat various age-related diseases [1].
Urolithin A, a metabolite of ellagic acid, demonstrates geroprotective and antioxidant properties across various model systems [2, 3]. Although it exhibits minimal uncoupling activity [4], it can induce mitophagy, likely through mechanisms unrelated to oxidative phosphorylation uncoupling [5]. We hypothesized that enhancing the uncoupling properties of urolithin A could augment its mitophagy induction and consequently boost its biological efficacy. Previous research has indicated that appending linear alkyl groups containing six carbon atoms to the molecule can enhance its uncoupling properties [6]. Building upon this, we synthesized a derivative of urolithin A with a C6H13 membranophilic group (designated as KIRS152). This compound exhibits stronger uncoupling activity and better cellular accumulation compared to the original urolithin A. Our aim was to compare the geroprotective effects of Urolithin A and KIRS152 in a model of cisplatin-induced aging in human skin fibroblasts.
Methods: The study involved normal primary fibroblasts from healthy donors. Cytotoxicity of urolithin A and KIRS152 on primary fibroblasts was assessed using the resazurin test. Subsequently, the impact of non-toxic concentrations of the compounds on cisplatin-induced senescence markers was evaluated. Fibroblasts were exposed to

10 μM cisplatin for 24 hours to induce senescence, followed by treatment with urolithin A and KIRS152 at concentrations ranging from 0.5 to 100 μM for six days. Senescent phenotype markers were analyzed on the seventh day.
The senescence index (SI) was determined as the main aging marker using a flow cytometer Amnis Flowsight with a visualization system. SI was calculated using the formula: $SI = ((nAF - 1) + 5 \times (nD - 1))/2$, where nAF represents normalized lipofuscin autofluorescence and nD represents normalized cell diameter. Lipofuscin accumulation and cell size increase with aging, hence higher SI values indicate older cells. Additionally, other methods used to assess senescence phenotype manifestation included histochemical assessment of senescence-associated β-galactosidase (SA-β-gal) and determination of GDF15 cytokine content (a marker of senescence-associated secretory phenotype) using enzyme immunoassay.
Results: Analysis of the resazurin test results showed that incubation of fibroblasts with urolithin A for 24 hours reduced fibroblast viability starting at a concentration of 20 μM. While KIRS152 is nontoxic up to 100 μM. Analysis of the results of SI evaluation on flow cytofluorimeter showed that both KIRS152 and urolithin A dose-dependently reduced SI value in the concentration range from 2 to 25 μM. Moreover, the geroprotective effect was more pronounced for KIRS152 than for urolithin A. Since the geroprotective effect of urolithin A is thought to be related to its ability to induce mitophagy, suppression of mitophagy should abolish this effect. To test this assumption, we used SBI-0206965, an inhibitor of ULK1 kinase, which plays a critical role in autophagy. Pretreatment of fibroblasts with SBI-0206965 abolished the geroprotective effect of KIRS152 and urolithin A. Which is in favor of the assumption that the mechanism of cellular action of the investigated substances may be related to the induction of mitophagy, which is a type of selective autophagy. To evaluate the effect of the geroprotectors KIRS152 and urolithin A on the manifestation of the senescence phenotype, the proportion of SA-β-gal-positive cells after histochemical staining was measured. SA-β-gal is a biomarker of activation of the lysosomal senescence pathway, for which a significant increase in expression with increasing cell age has been shown. Analysis of the data demonstrated a dose-dependent decrease in the proportion of SA-β-gal-positive cells for both KIRS152 and urolithin A over the concentration ranges investigated. Since KIRS152 markedly reduced IS but had a smaller effect on the proportion of SA-β-gal-positive cells, it can be assumed that it interferes with the accumulation of lipofuscin in the cell. Evaluation of the content of the growth and differentiation factor GDF15 in the incubation medium, the increase in the level of which is also a sign of senescence, showed that KIRS152 at a concentration of 2 μM prevents the cisplatin-induced increase in GDF15 secretion, while urolithin A has no such effect.
Conclusion: Both KIRS152 and urolithin A demonstrated dose-dependent reduction in SA-β-positive cells and SI in the cisplatin-induced aging model of human skin fibroblasts. KIRS152 exhibited a superior effect on SI compared to urolithin A at similar concentrations. Pretreatment with ULK1 kinase inhibitor abolished the geroprotective properties of both compounds, implicating mitophagy/autophagy in their protective mechanisms. Furthermore, only KIRS152 prevented enhanced GDF15 secretion and displayed lower toxicity to human fibroblasts compared to urolithin A, suggesting its superior geroprotective properties.
Funding: The study is supported by RNF (No. 23-14-00061).

Список литературы/References

1. Guilbaud E., Sarosiek K.A., Galluzzi L. Inflammation and mitophagy are mitochondrial checkpoints to aging. *Nat Commun*. 2024;15:3375. doi 10.1038/s41467-024-47840-1
2. Liu C.F., Li X.L., Zhang Z.L. et al. Antiaging Effects of Urolithin A on Replicative Senescent Human Skin Fibroblasts. *Rejuvenation Res*. 2019;22(3):191-200. doi 10.1089/rej.2018.2066
3. Jimenez-Loygorri J.I. et al. Mitophagy curtails cytosolic mtDNA-dependent activation of cGAS/STING inflammation during aging. *Nat Commun*. 2024;15:830
4. Esselun C., Theyssen E., Eckert G.P. Effects of Urolithin A on Mitochondrial Parameters in a Cellular Model of Early Alzheimer Disease. *Int J Mol Sci*. 2021;22:8333. doi 10.3390/ijms22158333
5. D'Amico D., Andreux P.A., Valdés P., Singh A., Rinsch C., Auwerx J. Impact of the Natural Compound Urolithin A on Health, Disease, and Aging. *Trends Mol Med*. 2021;27(7):687-699
6. Krasnov V.S., Kirsanov R.S., Khailova L.S. et al. Alkyl esters of 7-hydroxycoumarin-3-carboxylic acid as potent tissue-specific uncouplers of oxidative phosphorylation: Involvement of ATP/ADP translocase in mitochondrial uncoupling. *Arch Biochem Biophys*. 2022;728:109366. doi 10.1016/j.abb.2022.109366

doi 10.18699/bgrs2024-13-14

Применение qPCR для оценки эффективности удаления объемных повреждений ДНК *in vitro*

Попов А.*, Петрусева И., Лаврик О.

Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
** depolice@mail.ru*

Ключевые слова: репарация ДНК; объемные повреждения; PCR; эксцизионная репарация нуклеотидов; долголетие

Мотивация и цель: Белки эксцизионной репарации нуклеотидов (nucleotide excision repair, NER) удаляют из ДНК широкий спектр объемных повреждений, в том числе аддукты, формирующиеся под действием УФ-света и вредных полициклических соединений из окружающей среды. Оценка функционального статуса NER в клетках имеет важное значение как для фундаментальных исследований, так и при диагностике различных патологических состояний. В большинстве работ по изучению NER оценивали специфическую эксцизионную активность белков прямой детекцией продуктов эксцизии модельных ДНК либо путем постэксцизионного радиоактивного мечения 3'-концов этих продуктов [1–4]. Существенным недостатком этих методов, ограничивающим их применение в клинике, является необходимость использования радиоактивной метки при создании модельной ДНК либо детекции продуктов эксцизии. В связи с этим особую значимость для определения активности NER *in vitro* приобретают методы, основанные на количественной ПЦР (quantitative polymerase chain reaction, qPCR). Возможность оценки эффективности протекания репарации путем измерения флуоресценции в сочетании с простой и быстрой процедурой выполнения делает методы, основанные на qPCR, перспективными инструментами для применения не только в исследовательских целях, но и в медицине. Целью настоящей работы стала разработка метода оценки эффективности удаления объемных повреждений *in vitro* с использованием qPCR.

Методы и алгоритмы: С использованием процедуры, описанной в работе [1], были синтезированы модельные ДНК (160 п.н.), содержащие в своем составе ненуклеозидную вставку на основе N-[6-(5(6)-флуоресцеинилкарбамоил) гексаноил]-3-амино-1,2- пропандиола (nFlu) либо N-[6-(9-антраценилкарбамоил) гексаноил]-3-амино-1,2-пропандиола (nAnt). Комплементарные цепи обоих типов ДНК-дуплексов содержали вставку на основе тетраэтиленгликоля (TEG), с 5'-стороны которой была введена фосфотиоатная группа (PS). Также была синтезирована немодифицированная ДНК (nm/TEG-ДНК), имитирующая продукт репарации. NER-компетентные экстракты клеток голого землекопа (*Heterocephalus glaber*) и мыши (*Mus musculus*) были приготовлены по протоколу, представленному в работе [5].

Результаты: Результаты проведенной сравнительной оценки амплификации nFlu/TEG-ДНК и nm/TEG-ДНК подтвердили возможность использования ДНК-субстрата, сконструированного нами; значение порогового цикла C(t) для nm/TEG- и nFlu/TEG-ДНК составило 12.12 ± 0.29 и 19.28 ± 0.52 цикла соответственно; разница в значениях порогового цикла между nFlu/TEG- и nm/TEG-ДНК

(dC(t)) составила 7.17 ± 0.43 цикла. Таким образом, субстрат и продукт реакции специфической эксцизии различимы в ходе проведения PCR.
Используя сконструированный субстрат nFlu/TEG-ДНК, мы провели оценку эффективности удаления объемных повреждений ДНК *in vitro* белками клеток долгоживущего голого землекопа и короткоживущей мыши с помощью qPCR. Эффективность репарации nFlu/TEG-ДНК была более высокой на всем временном промежутке инкубации с белками экстракта клеток *H. glaber* и после 30 минут составила 55.67 ± 0.42 %, тогда как после 30 минут инкубации с белками экстракта клеток *M. musculus* эффективность репарации была значительно ниже и составила 35.17 ± 1.42 %. Инкубация nm/TEG-ДНК с белками экстрактов в течение 30 минут не приводила к заметному изменению в значениях C(t), что говорит об отсутствии значимого влияния неспецифического воздействия белков экстрактов в обоих случаях. Белки экстракта клеток *H. glaber* практически в 1.5 раза более эффективно удаляли объемное повреждение nFlu из модельного ДНК-субстрата, чем белки экстракта клеток *M. musculus*, что согласуется с ранее описанными результатами, полученными альтернативным методом [6].
Дополнительно оценена эффективность репарации nAnt/TEG-ДНК и nFlu/TEG-ДНК в реакции NER, катализируемой белками экстракта клеток CHO. Для этого 160-звенные nAnt/TEG- и nFlu/TEG-ДНК (16 нМ) выдерживали с белками экстракта клеток CHO (1.6 мг/мл) в течение 30 минут, после чего проводили анализ восстановленной ДНК с помощью qPCR. Эффективность репарации nAnt/TEG-ДНК была несколько ниже и после 30 минут инкубации с белками составила 28.72 ± 1.92 %, в то время как для nFlu/TEG-ДНК этот показатель составил 36.77 ± 1.68 %.
Выводы: Представленные результаты демонстрируют, что разработанный способ позволяет детектировать различия в эффективности удаления белками клеточных экстрактов объемных повреждений, отличающихся по своей химической структуре и субстратным свойствам. Таким образом, разработанный метод открывает перспективы для дальнейшего использования в исследованиях, направленных не только на изучение репарации ДНК в клетках долгоживущих млекопитающих, но и на поиск функциональной взаимосвязи различных систем репарации ДНК, участвующих в удалении объемных повреждений.
Финансирование: Исследование поддержано РНФ (№ 22-14-00112).

The use of qPCR to assess the efficiency of removal of bulky DNA damages *in vitro*

Popov A.*, Petruseva I., Lavrik O.
Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia
** depolice@mail.ru*

Key words: DNA repair; bulky damages; PCR; nucleotide excision repair; longevity

Motivation and Aim: Nucleotide excision repair (NER) proteins remove a wide range of bulky damages from DNA, including adducts formed by UV light and harmful polycyclic compounds from the environment. The assessment of the functional status of NER in cells is important both for fundamental research and for the diagnosis of various diseases. In most studies of NER, the specific excision activity of proteins was evaluated

by direct detection of excision products of model DNA or by postexcision radioactive labeling of the 3' ends of these products [1–4]. A significant disadvantage of these methods, which limits their use in the clinic, is the need to use a radioactive label when creating model DNA, or detecting excision products. In this regard, methods based on quantitative PCR (quantitative polymerase chain reaction, qPCR) are of particular importance for determining the activity of NER *in vitro*. The ability to evaluate the efficiency of repair by measuring fluorescence, combined with a simple and fast procedure, makes qPCR-based methods promising tools for use not only for research purposes, but also in medicine. The purpose of this work was to develop a method for assessing the efficiency of bulky damage removal *in vitro* using qPCR.

Methods and Algorithms: Using the procedure described in [1], model DNA (160 bp) containing a non-nucleoside insert based on N-[6-(5(6)- fluoresceinyl carbamoyl)-hexanoyl]-3-amino-1,2-propanediol (nFlu), or N-[6-(9-anthracenylcarbamoyl)-hexanoyl]-3-amino-1,2-propanediol (nAnt) were synthesized. The complementary chains of both types of DNA duplexes contained tetraethylene glycol (TEG) insert, from the 5' side of which a phosphothioate group (PS) was introduced. Unmodified DNA (nm/TEG-DNA) was also synthesized, simulating the repair product. NER-competent extracts of naked mole rat (*Heterocephalus glaber*) and mouse (*Mus musculus*) cells were prepared according to the protocol presented in [5].

Results: The results of a comparative assessment of the amplification of nFlu/TEG-DNA and nm/TEG-DNA confirmed the possibility of using the DNA substrate designed by us; the value of the threshold cycle C(t) for nm/TEG- and nFlu/TEG-DNA was 12.12 ± 0.29 and 19.28 ± 0.52 cycles, respectively; the difference in the values of the threshold cycle. The difference between nFlu/TEG- and nm/TEG-DNA (dC(t)) was 7.17 ± 0.43 cycles. Thus, the substrate and the product of the specific excision reaction are distinguishable during PCR.

Using the constructed nFlu/TEG-DNA substrate, we evaluated the efficiency of removal of bulky DNA damage *in vitro* by proteins from cells of a long-lived naked mole rat and a short-lived mouse using qPCR. The efficiency of nFlu/TEG DNA repair was higher during the entire incubation period with *H. glaber* cell extract proteins and after 30 minutes was 55.67 ± 0.42 %, while after 30 minutes of incubation with *M. musculus* cell extract proteins, the repair efficiency was significantly lower and amounted to 35.17 ± 1.42 %. Incubation of nm/TEG DNA with extract proteins for 30 minutes did not lead to a noticeable change in C(t) values, which indicates that there was no significant effect of the nonspecific effect of extract proteins in both cases. The proteins of the *H. glaber* cell extract were almost 1.5 times more effective at removing bulky nFlu damage from the model DNA substrate than the proteins of the *M. musculus* cell extract, which is consistent with the previously described results obtained by an alternative method [6].

The efficiency of nAnt/TEG-DNA and nFlu/TEG-DNA repair in the NER reaction catalyzed by proteins of the CHO cell extract was additionally evaluated. To do this, 160-link nAnt/TEG and nFlu/TEG DNA (16 nM) were incubated with CHO cell extract proteins (1.6 mg/ml) for 30 minutes, after which the recovered DNA was analyzed using qPCR. The efficiency of nAnt/TEG DNA repair was slightly lower, and after 30 minutes of incubation with proteins was 28.72 ± 1.92 %, while for nFlu/TEG DNA this indicator was 36.77 ± 1.68 %.

Conclusion: The presented results demonstrate that the developed method makes it possible to detect differences in the efficiency of removing bulky damages by proteins

of cellular extracts that differ in their chemical structure and substrate properties. Thus, the developed method opens up prospects for further use in research aimed not only at studying DNA repair in long-lived mammalian cells, but also at searching for the functional relationship of various DNA repair systems involved in the removal of bulky damages.

Funding: The study is supported RSF (No. 22-14-00112).

Список литературы/References

1. Evdokimov A. et al. New synthetic substrates of mammalian nucleotide excision repair system. *Nucleic Acids Res*. 2013;41:e123. doi 10.1093/nar/gkt301
2. Liu Z. et al. Resistance to nucleotide excision repair of bulky guanine adducts opposite abasic sites in DNA duplexes and relationships between structure and function. *PLoS One*. 2015;10:e0142068. doi 10.1371/journal.pone.0137124
3. Lukyanchikova N.V. et al. DNA with damage in both strands as affinity probes and nucleotide excision repair substrates. *Biochemistry (Moscow)*. 2016;81:263-274. doi 10.1134/S0006297916030093
4. Naumenko N., Petruseva I., Lomzov A., Lavrik O. Recognition and removal of clustered DNA lesions via nucleotide excision repair. *DNA Repair (Amst)*. 2021;108:103225. doi 10.1016/j.dnarep.2021.103225
5. Reardon J.T., Sancar A. Purification and characterization of Escherichia coli and human nucleotide excision repair enzyme systems. *Methods Enzymol*. 2006;408:189-213. doi 10.1016/S0076-6879(06)08012-8
6. Evdokimov A. et al. Naked mole rat cells display more efficient excision repair than mouse cells. *Aging (Albany NY)*. 2018;10:1454-1473. doi 10.18632/aging.101482

doi 10.18699/bgrs2024-13-15

Генетическое и фармакологическое подавление piwi модулирует продолжительность жизни и радиоустойчивость *Drosophila melanogaster*

Прошкина Е.*, Уляшева Н., Пакшина Н., Коваль Л., Юшкова Е.,
Соловьёв И., Шапошников М., Москалев А.
Институт биологии ФИЦ КомиНЦ УрО РАН, Сыктывкар, Россия
** proshkina.e.n@ib.komisc.ru*

Ключевые слова: PIWI; Argonaute; продолжительность жизни; радиоустойчивость; γ-излучение; *Drosophila melanogaster*

Мотивация и цель: Белки подсемейства PIWI семейства Argonaute и взаимодействующие с ними piwiРНК играют решающую роль в подавлении активности мобильных генетических элементов (МГЭ), которые являются одной из эндогенных причин геномной нестабильности. С возрастом происходит снижение эффективности данного механизма, в результате чего изменяется активность МГЭ и частота транспозиций, нарушаются структура и стабильность генома. Такие изменения связаны со старением и развитием возрастных заболеваний. В то же время стрессовые факторы, включая ионизирующие излучения, могут влиять на баланс активности МГЭ и усиливать возрастные изменения [1].

Как правило, функции белков PIWI и piwiРНК описывают в связи с поддержанием клеток зародышевой линии и половых желез. Однако в настоящее время расширяется понимание их функций в соматических тканях и регуляции продолжительности жизни (ПЖ) организма. Например, piwi подавляет возрастную экспрессию транспозонов в стволовых клетках кишечника дрозофилы и поддерживает эпителиальный гомеостаз [2]. Мутация гена *piwi* вызывает истощение жирового тела, нарушение липидного обмена и иммунитета, увеличение уровня повреждений ДНК у мух [3]. Также у дрозофилиной модели таупатии показано снижение активности piwi и дифференциальные изменения экспрессии piwiРНК в головах [4]. В то же время установлены как отрицательные, так и положительные эффекты мутаций *piwi* на ПЖ дрозофилы [3, 5].

В данной работе изучен вклад активности гена *piwi* в различных тканях имаго в регуляцию ПЖ и радиоустойчивости *Drosophila melanogaster*. Кроме того, проведен поиск потенциальных препаратов, способных модулировать активность белков PIWI, и оценен их геропротекторный и радиопротекторный потенциал.

Методы и алгоритмы: Исследование проводили на модели плодовой мушки *Drosophila melanogaster*. Для анализа возрастной динамики экспрессии генов и оценки геропротекторных и радиопротекторных свойств потенциальных ингибиторов PIWI использовали линию дикого типа Canton-S. В экспериментах по изучению влияния нокдауна гена *piwi* на ПЖ и радиоустойчивость особей получали путем скрещивания линий с UAS- и GAL4- конструкциями. UAS-линия экспрессирует двуцепочечную РНК для РНК-интерференции *piwi* под контролем

промотора UAS. GAL4-линии несут кондиционный (мифепристон-индуцируемый) драйвер GAL4-GeneSwitch в определенных тканях (*elav* – в нервной системе, *S106* – в жировом теле, *TIGS-2* – в кишечнике, *Mhc* – в мышцах). Полученные в результате скрещивания самцы и девственные самки были помещены в пробирки со средой, содержащей индуктор кондиционного GAL4 мифепристон (RU486, Sigma-Aldrich, США). Контрольные особи были получены в результате такого же скрещивания, но содержались на среде без мифепристона.

При анализе ПЖ мух содержали в стандартных условиях [6] до конца жизни. При изучении устойчивости к γ-изучению (120, 700 или 800 Гр) дрозофилы 15 сут находились в условиях, аналогичных экспериментам на ПЖ, после чего подвергались воздействию ионизирующего излучения и были помещены в пробирки без мифепристона, этанола или потенциальных ингибиторов PIWI. На основе полученных данных строили кривые дожития, рассчитывали медианную, среднюю ПЖ и возраст 90 % смертности. Статистическую значимость различий оценивали с помощью непараметрических критериев Колмогорова–Смирнова, Мантеля–Кокса, Гехана–Бреслоу–Вилкоксона, метода Ванг–Аллисона.

Анализ экспрессии генов с помощью метода ПЦР «в реальном времени» проводили на 15-е сутки жизни имаго с использованием целых дрозофил либо их частей (головы, тораксы или брюшки). Выделение РНК, проведение обратной транскрипции и постановку ПЦР выполняли с использованием реактивов и по методикам, описанным ранее [6]. Для оценки статистической значимости различий использовали критерий Манна–Уитни.

Компьютерный поиск потенциальных лигандов piwi дрозофилы и PIWIL1 человека, моделирование межмолекулярных взаимодействий и подбор концентраций проводили с помощью онлайн-инструментов DockThor (https://www.dockthor.lncc.br/v2) и AmDock (https://github.com/Valdes-Tresanco-MS/AMDock).

Результаты: Установлено, что с возрастом происходит одновременное повышение активности ретротранспозонов (в 2–12 раз, $p < 0.05$) и гена *piwi* (в 2–7 раз, $p < 0.05$) в головах, брюшках и во всем теле дрозофил. Это может указывать как на дерегуляцию биогенеза piwiРНК, так и на компенсаторный ответ на возрастающую с возрастом активность МГЭ и нарушение структуры хроматина.

В то же время нокдаун гена *piwi* в жировом теле и нервной системе увеличивал медианную ПЖ на 2–12 % ($p < 0.05$) по сравнению с контрольными особями. Но подавление активности *piwi* в кишечнике и мышцах, напротив, укорачивало жизнь мух. Как правило, повышенная ПЖ сопровождается усилением устойчивости организма к повреждающим факторам, таким как γ-излучение. Однако у дрозофил с нокдауном *piwi* эта связь была неоднозначна. У особей со сниженной активностью *piwi* в жировом теле наблюдали повышенную устойчивость к γ-излучению в дозе 700 Гр, тогда как нейрональный нокдаун этого гена увеличивал радиорезистентность мух. Кроме того, нокдаун *piwi* в нервной системе и в жировом теле повышал активность ретротранспозонов в 2–7 раз ($p < 0.05$) в головах и брюшках соответственно. Вместе с этим у них была повышена экспрессия генов антиоксидантной защиты (*Sod1*, *Prx5*), ответа на повреждение ДНК (*Gadd45*, *spn-B*, *Ku80*) и протеостаза (*Hsp27*, *Hsp68*, *Atg1*) до 8 раз ($p < 0.05$).

Мы предположили, что фармакологическое ингибирование PIWI также может оказывать влияние на ПЖ и радиоустойчивость организма. Проведенный компьютерный поиск позволил составить список фармакологических препаратов,

которые потенциально могут связываться с белком piwi дрозофилы и PIWIL1 человека. Среди данных соединений были выбраны пять веществ для скрининга геропротекторных свойств на модели дрозофилы – метопролол, атенолол, мизопростол, имипенем и финголимод. Для них был подобран спектр концентраций 1–5000 мкмоль/л, основываясь на компьютерном анализе и данных литературы. Все изученные препараты оказали положительное влияние на показатели ПЖ организма, но в некоторых случаях этот эффект присутствовал только при определенных концентрациях. Наиболее выраженное и воспроизводимое увеличение ПЖ вызывали метопролол в концентрациях 1–5000 мкмоль/л (3–9 %, $p < 0.0001$) и имипенем в концентрациях 1–100 мкмоль/л (4–14 %, $p < 0.0001$) у самок. В то же время потенциальные ингибиторы PIWI преимущественно снижали устойчивость дрозофил к воздействию γ-излучения. Защитное действие к облучению в дозе 120 Гр оказывали β-блокаторы атенолол и метопролол. Они повышали выживаемость дрозофил на 5–13 % ($p < 0.001$). Финголимод повышал выживаемость самок дрозофил после облучения в дозе 800 Гр на 18–32 % ($p < 0.001$). Кроме того, изученные препараты вызывали дифференциальные изменения активности генов антиоксидантной защиты (*Sod1*), ответа на повреждение ДНК (*Gadd45*, *mus210*, *Brca2*, *Ku80*) и протеостаза (*Hsp27*, *Hsp68*, *Atg1*, *Atg5*, *Ire*) до 5 раз ($p < 0.05$).

Выводы: Мы впервые показали, что нокдаун гена *piwi* в нервной системе и жировом теле вызывает увеличение ПЖ дрозофил и повышает активность генов, связанных со стресс-ответом и долголетием. При этом подавление *piwi* в жировом теле стимулирует устойчивость организма к γ-излучению, а нейрональный нокдаун снижает ее. Нами идентифицированы потенциальные ингибиторы белков PIWI. Для пяти из них (атенолола, метопролола, имипенема, финголимода и мизопростола) показаны геропротекторные свойства. Наиболее выраженное и воспроизводимое увеличение ПЖ вызывают метопролол и имипенем у самок. Влияние потенциальных ингибиторов PIWI на радиоустойчивость неоднозначно, зависит от концентрации препарата и дозы облучения.

Финансирование: Исследования выполнены в рамках государственного задания по теме «Генетические и функциональные исследования эффектов геропротекторных интервенций на модели *Drosophila melanogaster*» (№ 122040600022-1).

Genetic and pharmacological piwi suppression modulates lifespan and radioresistance of *Drosophila melanogaster*

Proshkina E.*, Ulyasheva N., Pakshina N., Koval L., Yushkova E., Solovyov I., Shaposhnikov M., Moskalev A.

Institute of Biology, FRC Komi SC UB RAS, Syktyvkar, Russia

** proshkina.e.n@ib.komisc.ru*

Key words: PIWI; Argonaute; lifespan; radioresistance; γ-radiation; *Drosophila melanogaster*

Motivation and Aim: Proteins of the PIWI subfamily of the Argonaute family and their interacting piwiRNAs play a crucial role in suppressing the activity of transposable elements (TEs), which are one of the endogenous causes of genomic instability. With

age, the effectiveness of this mechanism decreases, as a result of which the activity of TEs and the frequency of transpositions change, and the genome structure and stability is disrupted. Such changes are associated with aging and the development of age-related diseases. At the same time, stress factors, including ionizing radiation, can affect the balance of TE activity and enhance age-related changes [1].

Usually, the functions of PIWI proteins and piwiRNAs are described in relation to the maintenance of germline and gonad cells. However, the understanding of their functions in somatic tissues and the regulation of lifespan of an organism is currently expanding. For example, piwi suppresses age-related transposon expression in *Drosophila* intestinal stem cells and maintains epithelial homeostasis [2]. Mutation of the *piwi* gene causes depletion of the fat body, impaired lipid metabolism and immunity, and an increase in the level of DNA damage in flies [3]. Also, the *Drosophila* model of tauopathy showed a decrease in piwi activity and differential changes in piwiRNA expression in heads [4]. At the same time, both negative and positive effects of *piwi* mutations on the lifespan of *Drosophila* have been established [3, 5].

In this work, we studied the contribution of *piwi* gene activity in various imago tissues to the regulation of lifespan and radioresistance of *Drosophila melanogaster*. In addition, a search was carried out for potential drugs capable of modulating the activity of PIWI proteins, and their geroprotective and radioprotective potential was assessed.

Methods and Algorithms: The study was carried out on a model of the fruit fly *Drosophila melanogaster*. To analyze the age-related dynamics of gene expression and evaluate the geroprotective and radioprotective properties of potential PIWI inhibitors, the wild-type strain Canton-S was used. In experiments to study the effect of *piwi* gene knockdown on lifespan and radioresistance, individuals were obtained by crossing flies with UAS and GAL4 constructs. The UAS strain expresses double-stranded RNA for *piwi* RNA-interference under the control of the UAS promoter. GAL4 strains carry the conditional (mifepristone-inducible) GAL4-GeneSwitch driver in certain tissues (*elav* – in the nervous system, *S106* – in the fat body, *TIGS-2* – in the intestines, *Mhc* – in muscles). The resulting males and virgin females were placed in tubes containing medium with the conditional GAL4 inducer mifepristone (RU486, Sigma-Aldrich). Control individuals were obtained as a result of the same crossing, but were kept on medium without mifepristone.

When analyzing the lifespan of flies, they were kept under standard conditions [6] until the end of life. When studying resistance to γ-radiation (120, 700, or 800 Gy), *Drosophila* were kept for 15 days under conditions similar to the lifespan experiments, after which they were exposed to radiation and were placed in tubes without mifepristone, ethanol, or potential PIWI inhibitors. Based on the obtained data, survival curves were constructed; median, average lifespan, and age of 90 % mortality were calculated. The statistical significance of differences was assessed using nonparametric Kolmogorov–Smirnov, Mantel–Cox, Gehan–Breslow–Wilcoxon tests, and the Wang–Allison method.

Analysis of gene expression using the real-time PCR method was carried out on the 15th day of adult life using whole flies or their parts (head, thorax or abdomen). RNA isolation, reverse transcription, and PCR were performed using reagents and methods described previously [6]. To assess the statistical significance of differences, the Mann–Whitney test was used.

A computer search for potential ligands of *Drosophila* piwi and human PIWIL1, modeling of intermolecular interactions and selection of concentrations were carried out

using the online tools DockThor (https://www.dockthor.lncc.br/v2) and AmDock (https://github.com/ Valdes-Tresanco-MS/AMDock).

Results: It has been established that there is an age-related simultaneous increase in the activity of retrotransposons (2–12 times, $p < 0.05$) and the *piwi* gene (2–7 times, $p < 0.05$) in the heads, abdomens, and whole body of fruit flies. This may indicate both deregulation of piwiRNA biogenesis and a compensatory response to TE activity and disruption of chromatin structure that increase with age.

At the same time, knockdown of the *piwi* gene in the fat body and nervous system increased the median lifespan by 2–12 % ($p < 0.05$) compared to control individuals. But suppression of *piwi* activity in the intestines and muscles, on the contrary, shortened the lives of flies. As a rule, increased lifespan is accompanied by increased resistance to damaging factors, such as γ-radiation. However, in *piwi* knockdown fruit flies, this association was ambiguous. In individuals with reduced *piwi* activity in the fat body, increased resistance to γ-radiation at a dose of 700 Gy was observed, while neuronal knockdown of this gene increased the radioresistance of flies. In addition, knockdown of *piwi* in the nervous system and fat body increased retrotransposon activity by 2–7 times ($p < 0.05$) in the heads and abdomens, respectively. However, they had increased expression of genes for antioxidant defense (*Sod1*, *Prx5*), DNA damage response (*Gadd45*, *spn-B*, *Ku80*) and proteostasis (*Hsp27*, *Hsp68*, *Atg1*) up to 8 times ($p < 0.05$).

We hypothesized that pharmacological inhibition of PIWI may also have an effect on lifespan and radioresistance of an organism. A computer search allowed us to compile a list of pharmacological drugs that can potentially bind to the *Drosophila* piwi protein and human PIWIL1. Among these compounds, five substances were selected to screen for geroprotective properties in the *Drosophila* model – metoprolol, atenolol, misoprostol, imipenem, and fingolimod. A concentration range of 1–5000 μmol/l was selected for them, based on computer analysis and literature data. All the drugs studied had a positive effect on the lifespan, but in some cases this effect was present only at certain concentrations. The most pronounced and reproducible increase in lifespan was caused by metoprolol at concentrations of 1–5000 μmol/l (3–9 %, $p < 0.0001$) and imipenem at concentrations of 1–100 μmol/l (4–14 %, $p < 0.0001$) in females. At the same time, potential PIWI inhibitors predominantly reduced the resistance of *Drosophila* to γ radiation. The β-blockers atenolol and metoprolol had a protective effect against radiation at a dose of 120 Gy. They increased the survival rate of *Drosophila* by 5–13 ($p < 0.001$). Fingolimod increased the survival of *Drosophila* females after irradiation at a dose of 800 Gy by 18–32 % ($p < 0.001$). In addition, the studied drugs caused differential changes in the activity of antioxidant defense genes (*Sod1*), response to DNA damage (*Gadd45*, *mus210*, *Brca2*, *Ku80*) and proteostasis (*Hsp27*, *Hsp68*, *Atg1*, *Atg5*, *Ire*) up to 5 times ($p < 0.05$).

Conclusion: We have shown for the first time that knockdown of the *piwi* gene in the nervous system and fat body causes an increase in the lifespan of *Drosophila* and increases the activity of genes associated with stress-response and longevity. In this case, suppression of *piwi* in the fat body stimulates the resistance to γ-radiation, and neuronal knockdown reduces it. We have identified potential inhibitors of PIWI proteins. Five of them (atenolol, metoprolol, imipenem, fingolimod and misoprostol) have shown geroprotective properties. The most pronounced and reproducible increase in lifespan cause by metoprolol and imipenem in females. The effect of potential PIWI inhibitors on radioresistance is ambiguous and depends on the drug concentration and radiation dose.

Funding: The study was carried out within the framework of the state task on the theme "Genetic and functional studies of the effects of geroprotective interventions on the *Drosophila melanogaster* model" (No. 122040600022-1).

Список литературы/References

1. Yushkova E., Moskalev A. Transposable elements and their role in aging. *Ageing Res Rev*. 2023;86:101881. doi 10.1016/j.arr.2023.101881
2. Tang X. et al. Piwi maintains homeostasis in the Drosophila adult intestine. *Stem Cell Rep*. 2023;18(2):503-518. doi 10.1016/j.stemcr.2023.01.001
3. Jones B.C. et al. A somatic piRNA pathway in the Drosophila fat body ensures metabolic homeostasis and normal lifespan. *Nat Commun*. 2016;7:13856. doi 10.1038/ncomms13856
4. Sun W. et al. Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. *Nat Neurosci*. 2018;21(8):1038-1048. doi 10.1038/s41593-018-0194-1
5. Ma Z. et al. Epigenetic drift of H3K27me3 in aging links glycolysis to healthy longevity in Drosophila. *Elife*. 2018;7:e35368. doi 10.7554/eLife.35368
6. Proshkina E. et al. Tissue-specific knockdown of genes of the Argonaute family modulates lifespan and radioresistance in Drosophila melanogaster. *Int J Mol Sci*. 2021;22(5):2396. doi 10.3390/ijms22052396

doi 10.18699/bgrs2024-13-16

Эффекты репродуктивных технологий на поведение и развитие мозга у мышей линии B6.Cg-Tg, моделирующих болезнь Паркинсона

Рожкова И.[1]*, Козенева В.[1, 2], Брусенцев Е.[1], Рахманова Т.[1, 2], Шавшаева Н.[1, 2], Афанасова С.[1, 2], Лебедева Д.[1], Окотруб С.[1], Игонина Т.[1], Амстиславский С.[1]
[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*
[2] *Новосибирский государственный университет, Новосибирск, Россия*
** sirena23@yandex.ru*

Ключевые слова: болезнь Паркинсона; трансгенная модель; эмбриотрансфер; культивирование *in vitro*; отдаленные эффекты

Мотивация и цель: Болезнь Паркинсона (БП) является нейродегенеративным заболеванием человека. Для БП характерна прогрессирующая с возрастом ригидность мышц, тремор, брадикинезия, аномалии в нигростриарной системе головного мозга и альфа-синуклеинопатия [1–3]. Предпринимаются попытки установления взаимосвязи между применением вспомогательных репродуктивных технологий (ВРТ) и возникновением различных нейропатологий путем анализа клинических данных, однако выводы этих работ противоречивы [4–6], что создает мотивацию для постановки экспериментов на лабораторных животных. Трансгенные мыши линии B6.Cg-Tg(PrNp-SNCA*A53T)23Mkle (далее в тексте B6.Cg-Tg) имеют мутацию A53T в гене SNCA альфа-синуклеина человека. Ранее была оценена плотность нейронов в структурах мозга, которые играют ключевую роль в развитии БП, а также охарактеризовано поведение мышей B6.Cg-Tg [7]. Таким образом были выбраны маркеры, характеризующие мышей B6.Cg-Tg в качестве трансгенной модели БП. Влияние репродуктивных технологий на развитие болезни БП до настоящего момента не изучали. Целью данного исследования было изучение эффектов ВРТ на формирование фенотипа, характерного для БП у мышей B6.Cg-Tg. Задачами исследования было изучить, как культивирование *in vitro* и перенос эмбрионов влияют на поведение и развитие головного мозга у потомков.
Методы и алгоритмы: Исследования проводили на мышах, моделирующих БП – линия B6.Cg-Tg(PrNp-SNCA*A53T)23Mkle (далее в тексте B6.Cg-Tg), которые имеют мутацию A53T в гене SNCA альфа-синуклеина человека. В качестве контроля использовали мышей C57BL/6 дикого типа (wild type; далее в тексте WT), на основе которых получена линия B6.Cg-Tg. Животных содержали в стандартных условиях SPF-вивария ИЦиГ СО РАН. Все исследования были одобрены Комиссией по биоэтике ИЦиГ СО РАН (Протокол № 145 от 29 марта 2023 г). Самок C57BL/6 (WT) в возрасте 8–12 недель спаривали с гемизиготными по мутации A53T самцами мышей B6.Cg-Tg. Либо оставляли самок до родов, либо получали преимплантационные эмбрионы, культивировали их *in vitro* и трансплантировали псевдобеременным реципиентам, как описано нами ранее [8]. Потомков мужского пола, рожденных в результате естественного спаривания или применения репродуктивных технологий, генотипировали на наличие мутации A53T. В результате были сформированы следующие группы: 1) самцы C57BL/6, рожденные естественным путем (группа WT-CTL); 2) самцы B6.Cg-Tg, рожденные естественным

путем (группа B6.Cg-Tg-CTL); 3) самцы C57BL/6, рожденные с применением ВРТ (группа WT-ET – embryo transfer); 4) самцы B6.Cg-Tg, рожденные с применением ВРТ (группа B6.Cg-Tg-ET). У потомков в возрасте 6 месяцев исследовали поведение в тестах «открытое поле» (ОП), «приподнятый крестообразный лабиринт» (ПКЛ) и «Рота-род» (РР). Затем проводили интракардиальную перфузию для фиксации ткани головного мозга, изготавливали и окрашивали срезы для исследования черной субстанции (ЧС) согласно методу, описанному ранее [7, 8]. Плотность нейронов, меченных антителами против тирозингидроксилазы и против альфа-синуклеина, оценивали с использованием микроскопа Axio Imager.M2 (Carl Zeiss, Германия) с камерой Zeiss Axiocam 506 mono (Carl Zeiss, Германия). Анализ результатов поведенческих тестов проводили в программном пакете STATISTICA v. 12.0 (StatSoft, Inc., США). Данные были проверены на нормальность с помощью теста Шапиро–Уилка. Результаты поведенческих тестов оценивали при помощи двухфакторного ANOVA. Плотность нейронов оценивали с помощью одностороннего дисперсионного анализа Краскела–Уоллиса. Уровень значимости рассматривался как $p < 0.05$.

Результаты: Двухфакторный ANOVA выявил эффект «генотипа», но не «способа размножения» (с применением ВРТ или без него) на параметры тестов ОП и ПКЛ. Между тем в тесте РР было выявлено взаимодействие факторов «линии» и «способа размножения» [$F(1,27) = 5.658$; $p < 0.05$] по латентному времени на момент падения животного. У самцов WT-ET, рожденных после применения ВРТ, латентное время на момент падения животного было достоверно меньше, чем у самцов дикого типа в контрольной группе и линии B6.Cg-Tg в группе ЕТ. Различия по плотности дофаминовых нейронов и нейронов с альфа-синуклеином в ЧС представлены на рис. 1 и 2. Статистический анализ выявил у мышей WT-ET более низкую плотность дофаминовых нейронов ($p < 0.01$) по сравнению с группой WT-CTL (0.22×10^5 [0.16×10^5; 0.26×10^5] против 0.93×10^5 [0.84×10^5; 1.08×10^5] соответственно). Мыши B6.Cg-Tg-ET имели более высокую плотность нейронов с альфа-синуклеином ($p < 0.05$), чем животные из группы WT-ET (0.69×10^5 [0.59×10^5; 0.75×10^5] против 0.21×10^5 [0.16×10^5; 0.25×10^5] соответственно).

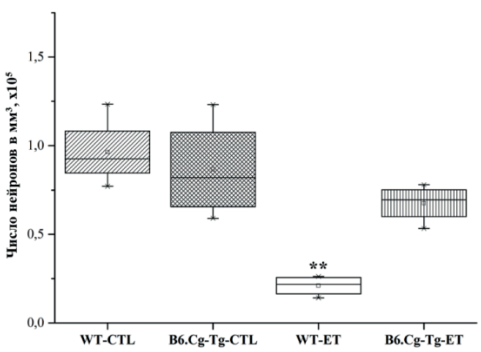


Рис. 1. Плотность дофаминовых нейронов в ЧС (substatia nigra – SN), нейроны мечены антителами против тирозингидроксилазы. ** $p < 0.01$ между группами WT-CTL и WT-ET

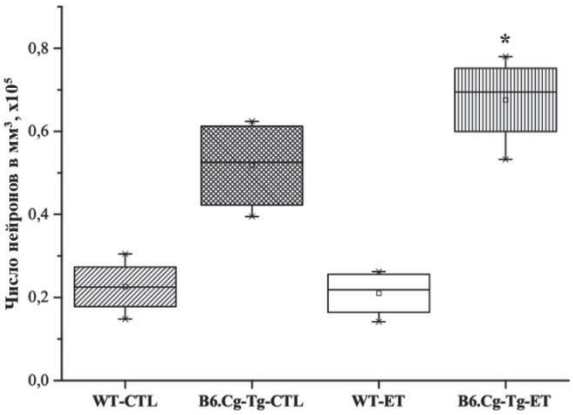


Рис. 2. Плотность нейронов с альфа-синуклеином в ЧС (substatia nigra – SN), нейроны мечены антителами против альфа-синуклеина. * $p < 0.05$ между группами B6.Cg-Tg-ET и WT-ET

Выводы: У потомков мышей дикого типа C57BL/6, полученных с применением ВРТ, наблюдали сниженную плотность дофаминовых нейронов в ЧС, а также снижение латентного времени до падения животного в тесте РР. У потомков мышей B6.Cg-Tg, полученных с применением ВРТ, наблюдали повышенную плотность нейронов с альфа-синуклеином в ЧС. Результаты данного исследования свидетельствуют о том, что репродуктивные технологии могут по-разному влиять на развитие потомков в зависимости от генотипа.
Финансирование: Исследование поддержано РНФ (проект № 23-25-00123).

Effects of reproductive technologies on behavior and brain development in B6.Cg-Tg mice modeling Parkinson's disease

Rozhkova I.[1]*, Kozeneva V.[1, 2], Brusentsev E.[1], Rakhmanova T.[1, 2], Shavshaeva N.[1, 2], Afanasova S.[1, 2], Lebedeva D.[1], Okotrub S.[1], Igonina T.[1], Amstislavsky S.[1]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *sirena23@yandex.ru*

Key words: Parkinson disease; transgenic model; embryo transfer; *in vitro* culture; long-term effects

Motivation and Aim: Parkinson's disease (PD) is a neurodegenerative human disease. PD is characterized by age-related muscle rigidity, tremor, bradykinesia, abnormalities in the nigrostriatal system of the brain and alpha-synucleinopathy [1–3]. Attempts to establish a relationship between the use of assisted reproductive technologies (ART) and the frequency and severity of various neuropathologies by analyzing clinical data led to contradictory conclusions [4–6], and motivate for experiments on the laboratory animal models. Transgenic mice of the B6.Cg-Tg(PrNp-SNCA*A53T)23Mkle/J strain (further referred to as B6.Cg-Tg) have the A53T mutation in the human alpha-synuclein SNCA

gene. Previously, the density of neurons in brain structures that play the key role in the development of PD was assessed, and the behavior of B6.Cg-Tg mice was characterized [7]. Thus, markers were selected that characterize B6.Cg-Tg mice as a transgenic model of PD. The effects of reproductive technologies on the development of PD disease have not been studied so far. The purpose of this study was to address the effects of ARTs on the development of PD phenotype in B6.Cg-Tg mice. The objectives of this study were to examine how *in vitro* culture and embryo transfer affect (1) behavior; (2) brain development in offspring.

Methods and Algorithms: Study was carried out on B6.Cg-Tg(PrNp-SNCA*A53T)23Mkle/J mouse strain modeling PD (referred further as B6.Cg-Tg) which has the A53T mutation in the human alpha-synuclein SNCA gene. Wild type C57BL/6 mice (referred further as WT) from which the B6.Cg-Tg strain was derived were used as controls. The animals were kept under standard conditions in the SPF-Vivarium of the Institute of Cytology and Genetics of the SB RAS. Study was approved by the Bioethics Committee of the Institute of Cytology and Genetics SB RAS (Protocol No. 145, March 29, 2023). C57BL/6 females, 8–12 weeks old, were mated to hemizygous A53T male B6.Cg-Tg mice. The mated females were either left until birth, or preimplantation embryos were obtained, cultured *in vitro* and transferred to pseudopregnant recipient as we described previously [8]. Male offspring born as a result of natural breeding or born after the use of reproductive technologies were genotyped for the presence of the A53T mutation. The following experimental groups were formed: (1) born naturally C57BL/6 males (WT-CTL group); (2) naturally born B6.Cg-Tg males (B6.Cg-Tg-CTL group); (3) C57BL/6 males born using ART (WT-ET – embryo transfer group); (4) B6.Cg-Tg males born using ART (group B6.Cg-Tg-ET). In offspring obtained at the age of 6 months, behavior was studied in the “open field” (OF), “elevated plus maze” (EPM), and “Rota-rod” (RR) tests. Thereafter intracardial perfusion was performed, brains were processed, brain slices were done and stained for the study of the substantia nigra (SN) according to the method described earlier [7, 8]. The density of neurons labeled with antibodies was assessed using an Axio Imager.M2 microscope (Carl Zeiss, Germany) with a Zeiss Axiocam 506 mono camera (Carl Zeiss, Germany). The results of behavioural tests were analyzed using the STATISTICA v. 12.0 software (StatSoft, Inc., USA). Data were tested for normality using the Shapiro–Wilk test. Results of behavioral tests were assessed using two-way ANOVA. Neuronal density was assessed using one-way Kruskal–Wallis analysis of variance. The significance level was considered $p < 0.05$.

Results: Two-factor ANOVA revealed the effect of the genotype, but not the method of reproduction (with or without ART) on the parameters of the OF and EPM tests. Meanwhile, the RR test revealed the interaction of the “genotype” and the “method of reproduction” factors [$F(1,27) = 5.658$; $p < 0.05$] for the latent time of the animal’s drop from the rod. This time for the males of the WT-ET males born with the help of ART was significantly less than for the control WT males and for the B6.Cg-Tg males born after ET. The densities of dopamine neurons and of neurons with alpha-synuclein in SN are shown in the Fig. 1, 2. Statistical analysis revealed a lower density of dopamine neurons in WT-ET mice ($p < 0.01$) compared to the WT-CTL mice (0.22×10^5 [$0.16 \times \times 10^5$; 0.26×10^5] vs. 0.93×10^5 [0.84×10^5; 1.08×10^5] respectively). B6.Cg-Tg-ET mice had a higher density of alpha-synuclein neurons ($p < 0.05$) compared with the animals from the WT-ET group (0.69×10^5 [0.59×10^5; 0.75×10^5] vs. 0.21×10^5 [$0.16 \times \times 10^5$; 0.25×10^5] respectively).

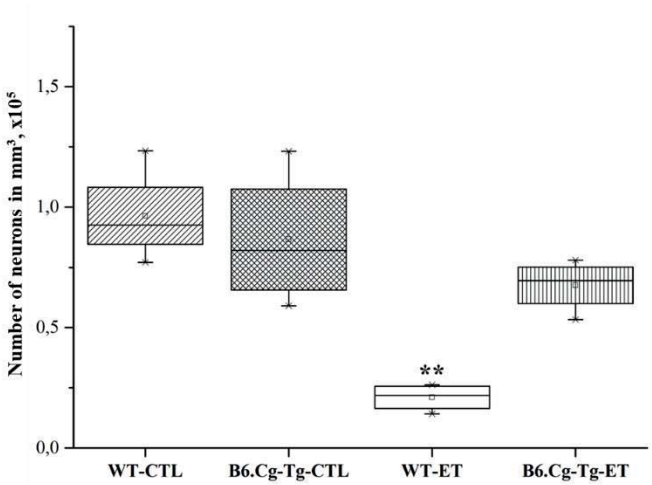


Fig. 1. Density of dopaminergic neurons in the substantia nigra (SN), neurons labeled with antibodies against tyrosine hydroxylase. ** $p < 0.01$ between WT-CTL and WT-ET groups

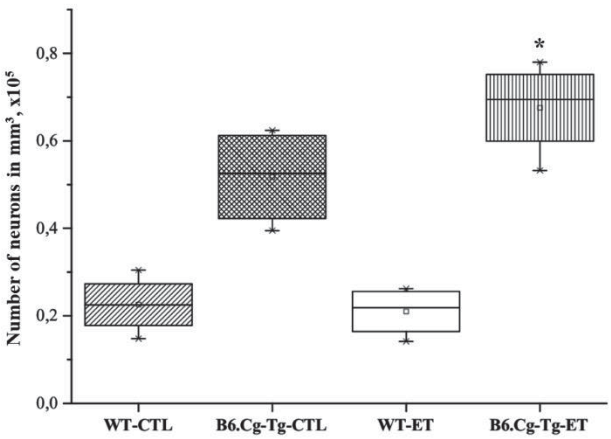


Fig. 2. Density of neurons with alpha-synuclein in substantia nigra (SN), neurons labeled with antibodies against alpha-synuclein. * $p < 0.05$ between B6.Cg-Tg-ET and WT-ET groups

Conclusion: In the offspring of wild-type C57BL/6 mice born with the use of ART, a reduced density of dopamine neurons in the SN was observed, as well as a decrease in the latent time of the animal's drop in the RR test. In the offspring of B6.Cg-Tg mice obtained with the use of ART, an increased density of neurons with alpha-synuclein in the SN was observed. The results of this study indicate that reproductive technologies may have different effects on the development of offspring depending on their genotype.
Funding: The study is supported Russian Science Foundation (No. 23-25-00123).

Список литературы/References

1. Halliday G.M., Del Tredici K., Braak H. Critical appraisal of brain pathology staging related to presymptomatic and symptomatic cases of sporadic Parkinson's disease. *J Neural Transmis Supp.* 2006;70:99
2. Beitz J.M. Parkinson's disease: a review. *Front Biosci.* 2014;6:65-74. doi 10.2741/S415
3. Dickson D.W. Neuropathology of Parkinson disease. Parkinsonism. *Relat Disord*. 2018;46:S30-S33. doi 10.1016/j.parkreldis.2017.07.033
4. Sandin S., Nygren K.G., Iliadou A., Hultman C.M., Reichenberg A. Autism and mental retardation among offspring born after *in vitro* fertilization. *Jama*. 2013;310(1):75-84
5. Andreadou M.T., Katsaras G.N., Talimtzi P. et al. Association of assisted reproductive technology with autism spectrum disorder in the offspring: An updated systematic review and meta-analysis. *Eur J Pediatr*. 2021;180(9):2741-2755
6. Smits R.M., Xavier M.J., Oud M.S. et al. *De novo* mutations in children born after medical assisted reproduction. *Hum Reprod.* 2022;37:1360-1369. doi 10.1093/humrep/deac068
7. Рожкова И.Н., Окотруб С.В., Брусенцев Е.Ю., Рахманова Т.А. и др. Анализ поведения и плотности нейронов в головном мозге мышей B6.Cg-Tg(Prnp-SNCA*A53T)23Mkle/J – модели болезни Паркинсона. *Российский физиологический журнал.* 2023;109:1199-1216
[Rozhkova I.N., Okotrub S.V., Brusentsev E.Y., Rakhmanova T.A. et al. Analysis of behavior and brain neuronal density in B6.Cg-Tg(Prnp-SNCA*A53T)23Mkle/J mice, a Parkinson's disease model. *J Evol Biochem Phys*. 2023;59:1633-1647. doi 10.1134/S0022093023050149]
8. Рожкова И.Н., Окотруб С.В., Брусенцев Е.Ю., Игонина Т.Н. и др. Эффекты вспомогательных репродуктивных технологий на социальное поведение мышей линии BTBR – модели расстройств аутистического спектра. *Российский физиологический журнал.* 2023;109:315-333
[Rozhkova I.N., Okotrub S.V., Brusentsev E.Y., Igonina T.N. et al. Effects of assisted reproductive technologies on social behavior of BTBR mice, a model of autism spectrum disorders. *J Evol Biochem Phys.* 2023;59:458-472. doi 10.1134/S0022093023020138]

doi 10.18699/bgrs2024-13-17

Функциональное значение полиморфизмов генов PLA2G12A, CFH, ARMS2 и C3, ассоциированных с развитием возрастной макулярной дегенерации

Румянцева Ю.В.*, Кожевникова О.С.
Институт цитологии и генетики СО РАН, Новосибирск, Россия
** Rumyantseva@bionet.nsc.ru*

Ключевые слова: SNP; возрастная макулярная дегенерация; система комплемента

Мотивация и цель: Возрастная макулярная дегенерация, наиболее распространенная причина потери зрения, представляет собой сложное многофакторное заболевание, в патогенезе которого играют роль старение, окружающая среда/образ жизни и генетические/эпигенетические факторы. Полногеномные ассоциативные исследования показали, что в патогенезе ВМД могут быть задействованы различные генетические локусы, содержащие гены трех основных путей: пути комплемента, метаболизма липидов и ремоделирования внеклеточного матрикса. Варианты этих генов могут взаимодействовать, и их эффект может модулироваться факторами риска ВМД.
Полиморфизмы rs10490924 (ARMS2), rs800292 (CFH) и rs6677604 (CFH) тесно связаны с экссудативной ВМД в российской популяции, а ответ на антиангиогенную терапию различался в зависимости от конкретного генотипа rs2285714 (CFI) и rs2230199 (C3) пациента [1].
Цель исследования – провести анализ in silico функционального значения полиморфных локусов генов PLA2G12A (rs2285714), CFH (rs800292, rs6677604), ARMS2 (rs10490924) и C3 (rs2230199), ассоциированных с ответом на анти-VEGF терапию ВМД.
Методы и алгоритмы: На основании баз данных (SIFT, SNPinfo, PolyPhen-2, HaploReg, GTExportal, RegulomeDB) проведена оценка функционального значения этих полиморфных локусов (несинонимические замены, эпигенетические эффекты, связь с экспрессией генов, ассоциации с альтернативным сплайсингом транскриптов генов).
Результаты: Среди пяти исследованных полиморфизмов rs2285714 (CFI) является синонимичной заменой, полиморфизм rs6677604 (CFH) вариантом интрона, а 3 SNP rs800292 (CFH), rs2230199 (C3) и rs10490924 (ARMS2) являются не синонимичными. Полиморфизмы rs80029 и rs2230199 не приводят к структурным изменениям белка и нарушениям его функции, а rs10490924 по данным PolyPhen-2 вероятно является вредоносным с показателем 0,994 (чувствительность 0,69, специфичность 0,97). По данным базы данных HaploReg, rs2285714 расположен в эволюционно консервативной области, rs2230199 — в области меток промоторных гистонов, а rs6677604 и rs2230199 — в области меток гистонов энхансера в различных тканях. Полиморфизмы rs6677604, rs10490924 и rs2230199 затрагивают сайты гиперчувствительные к ДНКазе1. Каждый из полиморфизмов затрагивает регионы регуляторных мотивов ДНК целого ряда тканей. Полиморфизм rs6677604 затрагивает регуляторные мотивы Ets_known1,

GATA_disc3, Hsf_known1, MZF1, RFX5_known1 и SPIB, а rs10490924 – AP-4_3, Ascl2, HEN1_2 и LBP-1_2. Также были определены SNP в сильном неравновесии по сцеплению с изученными полиморфизмами: от 6 SNP для rs2230199 до более 30 SNP для rs10490924. По данным базы данных SNPinfo, полиморфизмы rs2230199 C3 и rs2285714 CFI обладают наиболее значимым регуляторным потенциалом (0,33 и 0,36 соответственно). Те же SNP имеют самые высокие показатели вероятности RegulomeDB (0,86 и 0,6). Примечательно, что оба SNP расположены в областях энхансера экзонного сплайсинга (ESE) и сайленсера экзонного сплайсинга (ESS), а следовательно, могут влиять на процесс сплайсинга и изменять частоту альтернативного сплайсинга. Ранее нами было показано снижение уровня фактора FI, кодируемого геном CFI, во внутриглазной жидкости пациентов с ВМД [1]. Одной из причин данного эффекта может быть влиянием rs2285714 на тканеспецифический сплайсинг. По данным базы данных GTExportal, все пять SNP риска AMD в значительной степени включены в локусы экспрессии количественных признаков (eQTL), связанные с транскрипцией 18 генов-мишеней. rs10490924 может влиять на экспрессию трех генов (HTRA1, PLEKHA1 и ARMS2). Также rs10490924 относится к локусам сплайсинга количественных признаков (sQTL) для PLEKHA1 и ARMS2. rs2230199 имеет значение цис-eQTL, потенциально влияющее на экспрессию генов GPR108 и CTD-3128G10.7. rs2285714 имеет ассоциации тканеспецифичных транскриптов с 4 генами (CASP6, MCUB, CFI и PLA2GA12A) и представляет собой sQTL, имеющий большое значение для генов CFI и PLA2A12A.
rs6677604 может влиять на количество транскриптов мРНК 4 генов (CFHR1, CFHR2, CFHR3, ZBTB41) во многих тканях и органах и относится к sQTL для генов CFH и CFHR1. rs800292 может влиять на экспрессию 5 генов (CFH, CFHR1, CFHR3, CFHR4, KCNT2) и влияет на сплайсинг генов CFH, CFHR1, CFHR2 и F13B.
Выводы: Полиморфизмы ассоциированные с неоваскулярным ВМД в российской популяции и ответом на терапию обладают регуляторным потенциалом: участвуют в реализации эпигенетических механизмов, участвуют в контроле экспрессии генов и альтернативного сплайсинга.

Functional significance of polymorphisms of the PLA2G12A, CFH, ARMS2 and C3 genes, associated with the development of age-related macular degeneration

Rumyantseva Y.V.*, Kozhevnikova O.S.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** rumyantseva@bionet.nsc.ru*

Key words: SNP; age-related macular degeneration; complement

Motivation and Aim: Age-related macular degeneration the most common cause of vision loss, This complex, multifactorial disease, a major role in its pathogenesis play aging, different genetic/epigenetic factors, environment and lifestyle. Genome-wide association studies indicate that various genetic loci containing genes for three major pathways may be involved in the pathogenesis of AMD: the complement pathway, lipid

metabolism, and extracellular matrix remodeling. Variants of these genes may interact and their effect may modulate AMD risk factors. Polymorphisms rs10490924 (ARMS2), rs800292 (CFH) and rs6677604 (CFH) are closely associated with exudative AMD in the Russian population, and the response to antiangiogenic therapy differed depending on the specific genotype of rs2285714 (CFI) and rs2230199 (C3) of the patient [1].
Purpose of the study: To conduct an in silico analysis of the functional significance of the polymorphic loci of the PLA2G12A (rs (rs2285714), CFH (rs rs800292, rs6677604), ARMS2 (rs10490924) and C3 genes (rs2285714, rs800292, rs6677604, rs10490924 and rs2230199) associated with response to anti-VEGF therapy in AMD.
Methods and Algorithms: According to the databases (SIFT, SNPinfo, PolyPhen-2, HaploReg, GTExportal, RegulomeDB), we assessed the functional significance of these polymorphic loci (nonsynonymous substitutions, epigenetic effects, association with gene expression, associations with alternative splicing of gene transcripts).
Results: Among the five studied SNP, rs2285714 (CFI) is a synonymous substitution, the rs6677604 (CFH) polymorphism is an intron variant, and 3 SNPs rs800292 (CFH), rs2230199 (C3) and rs10490924 (ARMS2) are non-synonymous. Polymorphisms rs80029 and rs2230199 do not lead to structural changes in the protein or impairment of its function, and rs10490924, according to PolyPhen-2, is probably damaging with a score of 0.994 (sensitivity: 0.69; specificity: 0.97). According to the HaploReg database, rs2285714 is located in an evolutionarily conserved region, rs2230199 is located in the region of promoter histone marks region, and rs6677604 and rs2230199 are located in the region of enhancer histone marks in various tissues. The rs6677604, rs10490924 and rs2230199 polymorphisms affect DNase1 hypersensitive sites. Each of the polymorphisms affects regions of regulatory DNA motifs in a number of tissues. The rs6677604 polymorphism affects the regulatory motifs Ets_known1, GATA_disc3, Hsf_known1, MZF1, RFX5_known1, and SPIB, and the rs10490924 affects AP-4_3, Ascl2, HEN1_2, and LBP-1_2. SNPs in strong linkage disequilibrium with the studied polymorphisms were also identified: from 6 SNPs for rs2230199 to more than 30 SNPs for rs10490924. According to the SNPinfo database, the rs2230199 C3 and rs2285714 CFI polymorphisms have the most significant regulatory potential (0.33 and 0.36, respectively). The same SNPs had the highest RegulomeDB probability scores (0.86 and 0.6). Notably, both SNPs are located in the exonic splicing enhancer (ESE) and exonic splicing silencer (ESS) regions, and therefore can influence the splicing process and change the frequency of alternative splicing. Previously, we showed a decrease in the level of factor FI, encoded by the CFI gene, in the intraocular fluid of patients with AMD [1]. One of the reasons for this effect may be the influence of rs2285714 on tissue-specific splicing. According to the GTExportal database, all five AMD risk SNPs are significantly included in the expression quantitative trait loci (eQTL) associated with transcription of 18 target genes. rs10490924 might affect the expression of 3 genes (HTRA1, PLEKHA1 and ARMS2). Also the rs10490924 refers to splicing quantitative trait loci (sQTL) for PLEKHA1 and ARMS2 genes. rs2230199 has the cis-eQTL significance potentially affecting the expression of GPR108 and CTD-3128G10.7 genes. rs2285714 has the tissue-specific transcript associations with 4 genes (CASP6, MCUB, CFI and PLA2GA12A) and is sQTL with strong significance for CFI and PLA2A12A genes. rs6677604 can affect mRNA transcript abundance of 4 genes (CFHR1, CFHR2, CFHR3, ZBTB41) in many tissues and organs and refers to sQTL for CFH and CFHR1 genes. rs800292 might affect the expression of 5 genes (CFH, CFHR1, CFHR3, CFHR4, KCNT2) and cause sQTL for CFH, CFHR1, CFHR2 and F13B genes.

Conclusion: Polymorphisms associated with neovascular AMD in the Russian population and response to therapy have regulatory potential: they are involved in the implementation of epigenetic mechanisms, and participate in the control of gene expression and alternative splicing.

Список литературы/References

1. Kozhevnikova O.S., Fursova A.Z., Derbeneva A.S., Nikulich I.F., Tarasov M.S., Devyatkin V.A., Rumyantseva Y.V., Telegina D.V., Kolosova N.G. Association between Polymorphisms in CFH, ARMS2, CFI, and C3 Genes and Response to Anti-VEGF Treatment in Neovascular Age-Related Macular Degeneration. *Biomedicines*. 2022;10(7):1658. doi 10.3390/biomedicines10071658

doi 10.18699/bgrs2024-13-18

Изменение экспрессии транскрипционного фактора Stc в нервной системе и мышечной ткани увеличивает продолжительность здоровой жизни дрозофилы

Симоненко А.В.[1]*, Рощина Н.В.[1, 2], Пасюкова Е.Г.[1]

[1] *Научно-исследовательский центр «Курчатовский институт», Москва, Россия*

[2] *Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия*

* *symonenko@gmail.com*

Ключевые слова: дрозофила; *shuttle craft*; продолжительность жизни; подвижность; регуляция транскрипции; нервная система; нейрогенез; миозин; РНК-интерференция

Мотивация и цель: Увеличение здоровой продолжительности жизни является важнейшей целью исследований в биологии и медицине, имеющей значение для развития и подъема уровня благополучия человеческого общества в целом. Как показывают современные данные, если вплоть до возраста поздней зрелости долголетия можно достичь соблюдением здорового образа жизни, в пожилом возрасте решающую роль в определении качества и продолжительности жизни человека начинают играть генетические факторы.

В наших исследованиях генетического контроля продолжительности жизни основное внимание сосредоточено на белках-регуляторах транскрипции, поскольку именно они оказывают влияние сразу на многие гены, запуская в онтогенезе каскады генетической регуляции. Мы считаем важнейшими системами, определяющими продолжительность жизни организма, нервную, поскольку нейроны сохраняются в течение всего срока жизни и играют ключевую роль в регуляции основных функций организма, и мышечную, которая составляет основную часть массы тела (от 36 до 70 % у человека и 75 % у модельного объекта дрозофилы), внося важнейший вклад в общий метаболизм и реакцию организма на факторы внешней среды.

Мы исследуем влияние на продолжительность жизни уровня экспрессии в этих тканях гена *shuttle craft (stc)*, кодирующего транскрипционный фактор Stc, который влияет на развитие аксонов мотонейронов у дрозофилы [1].

Методы и алгоритмы: Для изменения экспрессии *stc* использовали бинарную систему Gal4-UAS [2]. Для индукции РНК-и нокдауна *stc* использовали линию с трансгеном, кодирующим двухцепочечную РНК *stc* (полученные из Vienna Drosophila Resource Centre линия VDRC_47973 и контрольная линия VDRC_60000), и линии драйверы, обеспечивающие паннейрональную экспрессию, экспрессию в центральной нервной системе эмбриона, мотонейронах и мышцах (полученные из Bloomington Drosophila Stock Center, линии BDSC_257750, BDSC_8165, BDSC_8816, BDSC_25756, соответственно). При необходимости за несколько поколений до использования в экспериментах линии были излечены от внутриклеточного паразита *Wolbachia*, влияющего на продолжительность жизни, развитие и метаболизм дрозофилы.

Уровень нокдауна *stc* у экспериментальных особей и его влияние на экспрессию возможных мишеней Stc проверяли при помощи кПЦР с использованием SYBR

Green I в системе MiniOpticon (Bio-Rad, США). Для оценки значимости отличий в количестве мРНК использовали процедуру ANOVA.

Продолжительности жизни измеряли у девственных мух одного пола, помещенных по 5 штук в пробирки со стандартным кормом, содержащим сахар, дрожжи, изюм и манную крупу, при температуре 25 °C и влажности 60 %, с чередованием света/темноты через 12 часов. Регистрацию погибших особей производили ежедневно, живых мух переносили на свежий корм еженедельно. Для оценки статистической значимости различий между кривыми выживания использовали тест Манна-Уитни, для оценки поздних эффектов – тест Флеминга-Харрингтона. Расчеты выполняли в программе OASIS 2 [3]. Для коррекции на множественность сравнений использовали поправку Бонферрони.

Измерение подвижности каждые десять дней проводили в пробирках диаметром 25 мм, содержащих по 5 девственных мух одного пола/возраста/генотипа, в горизонтальном положении помещенных в *Drosophila* Population Monitor (TriKinetics, США). Значимость отличий между средними значениями количества пересечений мухами трех светодиодных колец в течение пяти минут определяли, используя процедуру ANOVA.

Результаты: Учитывая, что у дрозофилы Stc был описан как нейрональный транскрипционный фактор [1], мы проанализировали влияние на продолжительность жизни паннейронального нокдауна *stc*. Такой нокдаун обусловил сокращение продолжительности жизни самцов на 13-15 % ($P<0.0001$). Мы исследовали последствия селективного нокдауна *stc* в эмбриональной нервной системе, и оказалось, что он увеличивает продолжительность жизни взрослых самцов на 12 % ($P<0.0001$). При этом подвижность таких самцов не отличалась от подвижности контрольных, а в возрасте сорок дней даже превышала ее ($P=0.004$), что является индикатором сохранения хорошего функционального состояния нервной и двигательной систем. Таким образом, можно предположить, что, поскольку в отличие от паннейронального нокдауна *stc*, селективный нокдаун в нервной системе на стадии эмбриогенеза полезен, то отрицательный вклад может вносить нокдаун *stc* на более поздних стадиях развития. Исходя из известной функции *stc* в развитии мотонейронов, мы определили влияние на продолжительность жизни селективного нокдауна в мотонейронах, и оказалось, что он не оказывал достоверного влияния на продолжительность жизни самцов ($P=0.795$). Вопрос о том, какие нейроны ответственны за снижение продолжительности жизни при паннейрональном нокдауне *stc*, остается пока открытым.

Для определения возможных мишеней транскрипционного фактора Stc, обусловливающих продление жизни, мы проанализировали изменения транскриптома в случае нокдауна *stc* в эмбриональной нервной системе и проверили их с помощью кПЦР. Среди выделенных нами мишеней оказались гены, определяющие развитие нервной системы (*N*, *Nej*, *Unc-115a*). Также, как возможные мишени Stc были определены несколько генов, участвующих в развитии мышечной системы, в частности, ген *Mhc*, кодирующий одну из субъединиц миозина. Учитывая этот факт, мы изучили влияние на продолжительность жизни нокдауна *stc* в миоцитах. Оказалось, что селективное снижение экспрессии *stc* в мышцах самцов увеличивает продолжительность их жизни на 19 % ($P<0.0001$).

Выводы: Мы показали, что селективный нокдаун нейронального транскрипционного фактора Stc в эмбриональной нервной системе дрозофилы обусловливает увеличение здоровой продолжительности жизни имаго с сохранением высокого функционального статуса нервной и мышечной систем. Также положительное влияние на продолжительность жизни оказывает снижение экспрессии Stc в мышечных клетках. Возможными мишенями Stc, опосредующими эти эффекты, является ряд генов, участвующих в развитии нервной системы, а также основной мышечный белок миозин.
Финансирование: Работа проведена в рамках выполнения государственного задания НИЦ «Курчатовский институт».

Changes in the expression of the Stc transcription factor in the nervous system and muscle tissue increase the healthspan of *Drosophila*

Symonenko A.V.[1]*, Roshina N.V.[1, 2], Pasyukova E.G.[1]

[1] *National Research Centre "Kurchatov Institute", Moscow, Russia*

[2] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

* *symonenko@gmail.com*

Key words: *Drosophila*; *shuttle craft*; lifespan; locomotion; transcription regulation; nervous system; neurogenesis; myosin; RNA interferention

Motivation and Aim: Increasing healthy lifespan is a crucial goal of research in biology and medicine, with significant implications for the development and enhancement of the well-being of human society as a whole. According to contemporary data, while longevity can be achieved up to late adulthood through a healthy lifestyle, genetic factors play a decisive role in determining the quality of life and lifespan in old age.

In our studies on the genetic control of lifespan, we focus primarily on transcriptional regulator proteins, as they influence multiple genes at once, triggering cascades of genetic regulation during ontogenesis. We consider the nervous system, which maintains neurons throughout life and plays a key role in regulating fundamental functions of the body, and the muscular system, which constitutes the majority of body mass (36–70 % in humans and 75 % in the model organism *Drosophila*), contributing significantly to overall metabolism and the organism's response to environmental factors, as the most critical systems determining lifespan.

We investigate the impact on lifespan of the expression level of the *shuttle craft* (*stc*) gene, which encodes the transcription factor Stc, known to affect the development of motor neuron axons in *Drosophila* [1], in these tissues.

Methods and Algorithms: To alter *stc* expression, we used the binary Gal4-UAS system [2]. For RNAi-induced *stc* knockdown, we utilized a transgenic line encoding double-stranded *stc* RNA (obtained from the Vienna Drosophila Resource Centre, line VDRC_47973 and control line VDRC_60000), and driver lines targeting pan-neuronal expression, expression in the embryonic central nervous system, motor neurons, and muscles (obtained from the Bloomington Drosophila Stock Center, lines BDSC_257750, BDSC_8165, BDSC_8816, BDSC_25756, respectively). If necessary, lines were treated

for the intracellular parasite *Wolbachia*, which affects lifespan, development, and metabolism of *Drosophila*, several generations before use in experiments.

The level of *stc* knockdown in experimental specimens and its effect on the expression of potential Stc targets were verified using qPCR with SYBR Green I in a MiniOpticon system (Bio-Rad, USA). The significance of differences in mRNA quantities was assessed using ANOVA.

Lifespan was measured in virgin flies of one sex, placed five per vial with standard medium containing sugar, yeast, raisins, and semolina, at a temperature of 25°C and humidity of 60 %, with a 12-hour light/dark cycle. The registration of dead individuals was carried out daily, and live flies were transferred to fresh food weekly. Statistical significance of differences between survival curves was assessed using the Mann-Whitney test, and late effects were evaluated using the Fleming-Harrington test. Calculations were performed in the OASIS 2 program [3]. Bonferroni correction was applied for multiple comparisons.

Locomotion was measured every ten days in 25 mm diameter tubes containing five virgin flies of the same sex/age/genotype, placed horizontally in a *Drosophila* Population Monitor (TriKinetics, USA). The significance of differences between mean values of the total fly crossings of three LED rings within five minutes was determined using ANOVA.

Results: Given that in *Drosophila*, Stc has been described as a neuronal transcription factor [1], we analyzed the impact of pan-neuronal *stc* knockdown on lifespan. This knockdown resulted in a 13–15 % reduction in male lifespan ($P < 0.0001$). We investigated the effects of selective *stc* knockdown in the embryonic nervous system and found that it increased the adult male lifespan by 12 % ($P < 0.0001$). Moreover, the mobility of these males was comparable to the controls, and at the age of forty days, it even surpassed that of the controls ($P = 0.004$), indicating the maintenance of good functional status of the nervous and motor systems. Thus, it can be hypothesized that, unlike the pan-neuronal knockdown, selective knockdown in the nervous system during the embryonic stage is beneficial, suggesting that *stc* knockdown specifically in adults may have a negative impact. Based on the known function of *stc* in motor neuron development, we examined the effect of selective knockdown in motor neurons and found that it did not significantly affect male lifespan ($P = 0.795$). The question of which neurons are responsible for the reduced lifespan with pan-neuronal *stc* knockdown remains still open.

To identify potential targets of the Stc transcription factor that might contribute to lifespan extension, we analyzed transcriptome changes resulting from *stc* knockdown in the embryonic nervous system and validated these findings using qPCR. Among the identified targets were genes involved in nervous system development (*N, Nej, Unc-115a*). Additionally, several genes involved in muscle system development, particularly the *Mhc* gene encoding one of the myosin subunits, were identified as possible Stc targets. Considering this, we studied the impact of *stc* knockdown in myocytes on lifespan. It was found that selective reduction of *stc* expression in male muscles increased their lifespan by 19 % ($P < 0.0001$).

Conclusion: We have shown that selective knockdown of the neuronal transcription factor Stc in the embryonic nervous system of *Drosophila* results in an increased healthspan of the imago while maintaining a high functional status of the nervous and muscular systems. Additionally, reducing the expression of Stc in muscle cells positively

affects lifespan. Potential targets of Stc mediating these effects include several genes involved in nervous system development, as well as the primary muscle protein myosin. *Funding*: The work was carried out within the state assignment of NRC “Kurchatov Institute”.

Список литературы/References

1. Stroumbakis N.D., Li Z., Tolias P.P. A homolog of human transcription factor NF-X1 encoded by the Drosophila *shuttle craft* gene is required in the embryonic central nervous system. *Mol Cell Biol.* 1996;16(1):192-201. doi 10.1128/MCB.16.1.192
2. Brand A.H., Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development.* 1993;118(2):401-415. doi 10.1242/dev.118.2.401
3. Han S.K., Lee D., Lee H. et al. OASIS 2: online application for survival analysis 2 with features for the analysis of maximal lifespan and healthspan in aging research. *Oncotarget*. 2016;7(35):56147-56152. doi 10.18632/oncotarget.11269

doi 10.18699/bgrs2024-13-19

Количественный метаболомный профиль сыворотки крови и мозга крыс OXYS – модели болезни Альцгеймера

Снытникова О.А.[1]*, Смоленцев А.А.[1], Телегина Д.В.[2], Центалович Ю.П.[1], Колосова Н.Г.[2]

[1] *Институт «Международный томографический центр» СО РАН, Новосибирск, Россия*

[2] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*

* *snytnikova_olga@tomo.nsc.ru*

Ключевые слова: Болезнь Альцгеймера; крысы OXYS; сыворотка; гиппокамп; ЯМР

Мотивация и цель: Метаболомный подход находит широкое применение в изучении патогенеза различных заболеваний, поскольку позволяет с высокой точностью оценивать характер изменения метаболизма при их развитии и выявлять потенциальные биомаркеры. Важным объектом в таких исследованиях являются животные – экспериментальные модели заболеваний человека, позволяющие получать знания, необходимые для разработки методов их диагностики, профилактики и лечения. Возраст – основной фактор риска Болезни Альцгеймера (БА), которая становится основной причиной сенильной деменции. Эффективных способов профилактики и лечения БА нет, что обусловлено неполнотой знаний патогенеза заболевания, которое десятилетиями развивается бессимптомно, а также дефицитом моделей самой распространенной (>95 %) спорадической формы БА. Настоящее исследование выполнено на преждевременно стареющих крысах OXYS – уникальной модели спорадической формы БА, у которых спонтанно развиваются все ключевые патогенетические и «клинические» признаки заболевания. Их последовательность: дисфункция митохондрий, гиперфосфорилирование тау-белка, нарушение длительной посттетанической потенциации, синаптическая недостаточность, деструктивные изменения нейронов, нарушения поведения и снижение когнитивных функций на ранних стадиях и их прогрессия на фоне повышения уровня APP, усиленного накопления Aβ и образования амилоидных бляшек в мозге, – соответствует современным представлениям о патогенезе спорадической формы БА у людей [1]. Цель работы – сравнение метаболома гиппокампа и сыворотки крови крыс OXYS на разных стадиях развития признаков БА для выявления наиболее перспективных с прогностической точки зрения биомаркеров – низкомолекулярных метаболитов, отражающих нарушения в метаболических циклах. Были исследованы крысы OXYS в «доклинический» период, предшествующий развитию признаков БА (в возрасте 20 дней), в период их манифестации (3–4 мес.) и активной прогрессии (16–18 мес.). Крысы линии Wistar того же возраста служили контролем.

Методы и алгоритмы: Методом метаболомного профилирования с использованием аналитической платформы на основе ЯМР спектроскопии проведен количественный анализ метаболитов мозга и крови крыс. Полученные данные проанализированы методами хемометрического и статистического анализа.

Результаты: Для метаболитов мозга (гиппокамп, 59 соединений) и сыворотки крови (55 соединений) были установлены диапазоны варьирования и средние

значения концентраций. Анализ полученных данных выявил основные закономерности изменения метаболома при старении, связанные прежде всего с энергетическим метаболизмом и нейротрансмиттерами. Установлено, что развитие, манифестация и прогрессия признаков БА у крыс OXYS происходит на фоне повышенного, по сравнению с крысами Wistar, уровня сцилло-инозитола и сниженного – гипотаурина в гиппокампе, которые были изменены у животных всех исследованных возрастных групп [2]. Полученные результаты позволили предположить, что повышенный уровень сцилло-инозитола в мозге может служить предиктором и потенциальным биомаркером развития признаков БА у крыс OXYS и, возможно, этого заболевания у людей. Однако в сыворотке крови уровень сцилло-инозитола достоверно снижался с возрастом и при этом не различался у крыс Wistar и OXYS, а уровень гипотаурина был ниже уровня детекции, что не позволяет использовать эти соединения в качестве доступных биомаркеров ранних стадий БА. Помимо сцилло-инозитола, было обнаружено еще 33 метаболита, общих для гиппокампа и сыворотки (рис. 1А), которые связаны с метаболизмом аминокислот, циклом трикарбоновых кислот, метаболизмом пирувата и азота. Следует отметить, что изменения уровня большинства метаболитов в гиппокампе и сыворотке крови в динамике развития признаков БА оказались реципрокными, что, скорее всего, обусловлено изменениями состояния гематоэнцефалического барьера на фоне развития нейродегенеративных процессов.

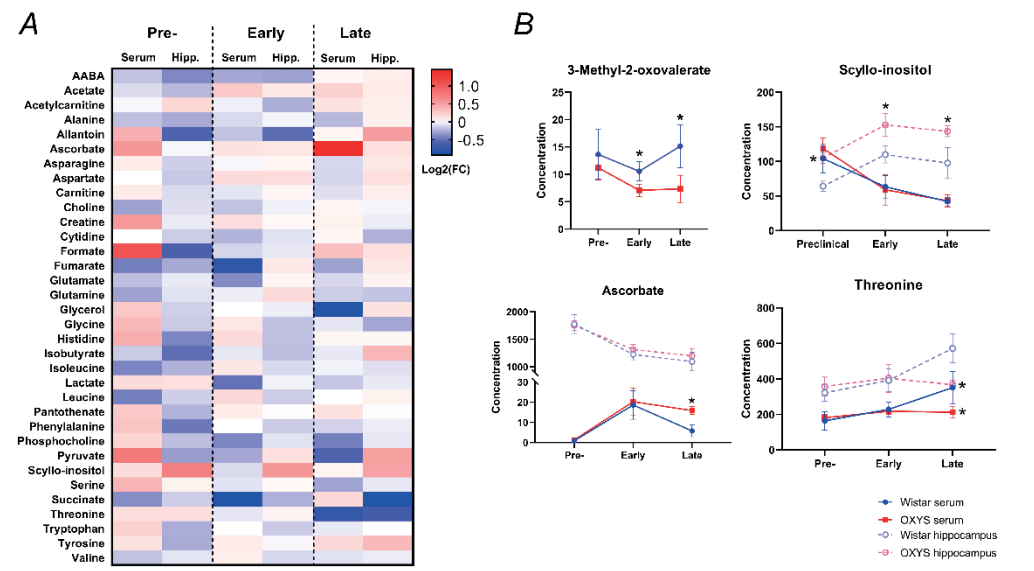


Рис. 1. А. Тепловая карта кратности изменения уровня общих для сыворотки и гиппокампа метаболитов между крысами Wistar и OXYS разного возраста (log2(FC)(Fold Change); В. Уровень 3-метил-2-оксовалериановой кислоты в сыворотке, сцилло-инозитола, аскорбата и треонина в сыворотке и гиппокампе крыс на доклинической (возраст 20 дней), стадии манифестации (3–4 мес.) и активной прогрессии заболевания (16–18 мес.) Данные представлены как среднее ± SD. * – достоверные различия между линиями

Согласно данным ЯМР-спектроскопии, на доклинической стадии заболевания (в возрасте 20 дней) в сыворотке крови крыс OXYS по сравнению с крысами Wistar достоверно изменялся уровень 9 метаболитов: аллантоина, креатина, глицина, гистидина, лизина и серина – повышался, а глутамина, изолейцина и лейцина – снижался. При этом в гиппокампе крыс в возрасте 20 дней уровни аллантоина и гистидина у крыс OXYS были достоверно ниже, чем у крыс Wistar, в то время как уровни других 6 общих метаболитов (лейцина, изолейцина, глутамина, креатина, глицина и серина) не изменялись. В период манифестации признаков БА (3–4 мес.) в сыворотке крыс OXYS были снижены уровни цитрата и 3-метил-2-оксовалериановой кислоты (одного из промежуточных метаболитов синтеза аминокислот с расщепленной цепью), ЯМР-сигналы которых не были обнаружены в образцах гиппокампа животных. Важно отметить, что в период выраженных нейродегенеративных изменений (18 мес.) уровень 3-метил-2-оксовалериановой кислоты в сыворотке крыс OXYS также был ниже, чем у крыс Wistar, что позволяет рассматривать этот метаболит как потенциальный биомаркер ранней и поздних стадий БА. Кроме того, в возрасте 18 месяцев в сыворотке крыс OXYS наблюдается повышение уровня аскорбата и снижение 2-кетоизовалерата, глицерола, кетолейцина, метионина и треонина. В гиппокампе 18-месячных крыс OXYS уровень треонина был, как и в сыворотке, достоверно ниже, чем у крыс Wistar, а уровни аскорбата и глицерола не различались (рис. 1В).
Выводы: Развитию признаков БА у крыс OXYS предшествует и сопутствует повышение уровня сцилло-инозитола и снижение – гипотаурина в гиппокампе. Сравнение метаболического профиля гиппокампа и сыворотки крыс выявило наличие 34 общих метаболитов, при этом направленность изменений с возрастом большинства из них была реципрокной в крови и мозге животных. На ранней и поздней стадиях развития БА-подобной патологии в сыворотке крови крыс OXYS снижен уровень 3-метил-2-оксовалериановой кислоты, что позволяет рассматривать этот метаболит как потенциальный доступный биомаркер заболевания.
Финансирование: Исследование поддержано проектом государственного бюджета № FWNR-2022-0016.

Quantitative metabolomic profile of blood serum and brain of OXYS rats – a model of Alzheimer's disease

Snytnikova O.A.[1]*, Smolentsev A.A.[1], Telegina D.V.[2], Tsentalovich Yu.P.[1], Kolosova N.G.[2]

[1] *International Tomography Center, SB RAS, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *snytnikova_olga@tomo.nsc.ru*

Key words: Alzheimer's disease; OXYS rats; hippocampus; serum; NMR

Motivation and Aim: The metabolomic approach is widely used in studying the pathogenesis of various diseases, since it allows one to accurately assess the nature of metabolic changes during their development and identify potential biomarkers. An important objects in such researches are animals – experimental models of human

diseases, which provide knowledge necessary for the development of methods for their diagnosis, prevention and treatment. Age is a major risk factor for Alzheimer's disease (AD), which becomes the leading cause of senile dementia. There are no effective methods for the prevention and treatment of AD, which is due to incomplete knowledge of the pathogenesis of the disease. Study of early stages of AD are difficult because AD develops asymptomatically for decades, and due to a lack of models for the most common (>95 %) sporadic form. The present study was performed on prematurely aging OXYS rats, a unique model of sporadic AD, in which all the key pathogenetic and "clinical" signs of the disease spontaneously develop. Their sequence (mitochondrial dysfunction, hyperphosphorylation of tau protein, impaired long-term post-tetanic potentiation, synaptic failure, destructive neuronal changes, behavioral disorders and cognitive decline in the early stages and their progression against the background of increased APP levels, increased accumulation of Aβ and the formation of amyloid plaques in the brain) corresponds to modern ideas about the pathogenesis of the sporadic form of AD in humans [1]. The purpose of the work is to compare the metabolome of the hippocampus and the serum of OXYS rats at different stages of development of signs of AD to identify the most promising biomarkers from a prognostic point of view – low molecular weight metabolites that reflect disturbances in metabolic cycles. OXYS rats were studied in the "preclinical" period preceding the development of signs of AD (at the age of 20 days), during their manifestation (3-4 months) and active progression (16-18 months). Age-matched Wistar rats were used as controls.

Methods and Algorithms: A quantitative analysis of metabolites in the brain and blood of rats was carried out using the method of metabolomic profiling using an analytical platform based on NMR spectroscopy. The obtained data were analyzed using chemometric and statistical analysis methods.

Results: For brain (hippocampus 59 compounds, cortex 44 compounds) and blood serum (55 compounds) metabolites, ranges of variation and mean concentrations were established. Analysis of the data obtained made it possible to identify the main metabolic patterns during aging, involved in the pathways of energy metabolism and metabolic changes in neurotransmitters. Analysis of the data obtained for brain metabolites revealed the main patterns of changes in the metabolome during aging, primarily associated with energy metabolism and neurotransmitters. It has been established that the development, manifestation and progression of signs of AD in OXYS rats occurs against the background of increased, compared to Wistar rats, levels of scyllo-inositol and decreased levels of hypotaurine in the hippocampus, which changed in animals of all age groups studied [2]. The findings suggest that elevated levels of scyllo-inositol in the brain may serve as a predictor and potential biomarker for the development of signs of AD in OXYS rats and, possibly, AD development in humans. However, in the blood serum, the level of scyllo-inositol significantly decreased with age and did not differ between Wistar and OXYS rats, and the level of hypotaurine was below the detection level, which does not allow the use of these compounds as available biomarkers of the early stages of AD. In addition to scyllo-inositol, 33 other metabolites were found to be common to the hippocampus and serum (Fig. 1A), which are associated with amino acid metabolism, the tricarboxylic acid cycle, pyruvate and nitrogen metabolism. It should be noted that changes in the level of most metabolites in the hippocampus and serum in the dynamics of the development of AD signs turned out to be reciprocal, which is most likely due to changes in the state of the blood-brain barrier against the background of the development of neurodegenerative processes.

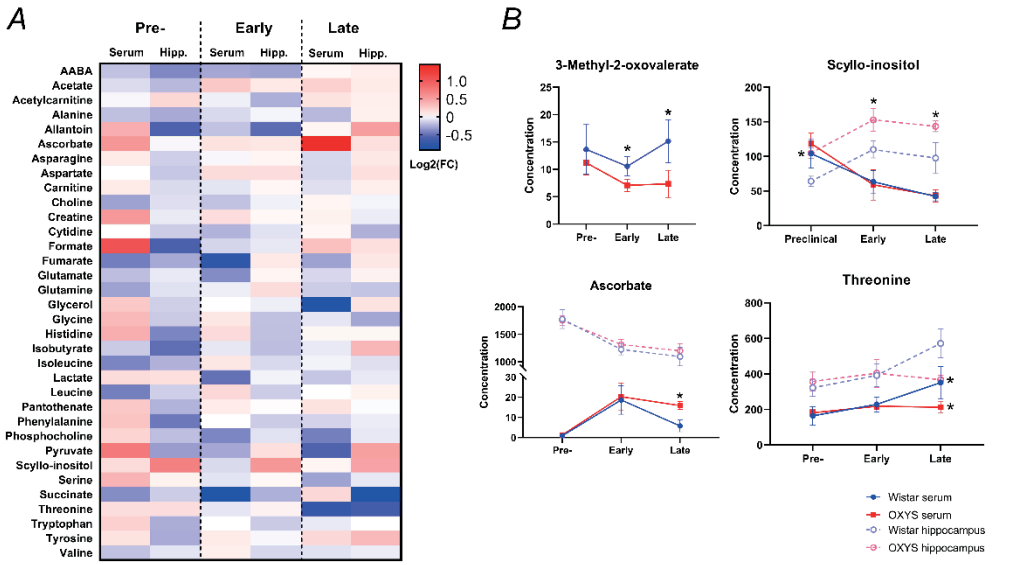


Fig. 1. A. Heat map of log2(FC) fold change in serum and hippocampal common metabolites between Wistar and OXYS rats of different ages; B) Levels of 3-methyl-2-oxovaleric acid in serum, scyllo-inositol, ascorbate and threonine in the serum and hippocampus of rats at the preclinical (age 20 days), stage of manifestation (3-4 months) and active progression of the disease (18 months). Data are presented as mean±SD. * – significant differences between lines

According to NMR spectroscopy data, at the preclinical stage of the disease (at the age of 20 days) in the serum of OXYS rats compared to Wistar rats, the level of 9 metabolites significantly changed: allantoin, creatine, glycine, histidine, lysine and serine increased, while glutamine, isoleucine and leucine decreased. In the hippocampus of rats at the age of 20 days, the levels of allantoin and histidine in OXYS rats were significantly lower than in Wistar rats, while the levels of other 6 common metabolites (leucine, isoleucine, glutamine, creatine, glycine and serine) have not changed. During the period of manifestation of signs of AD (3-4 months), the levels of citrate and 3-methyl-2-oxovaleric acid (one of the intermediate metabolites of the synthesis of branch-chain amino acids) were reduced in the serum of OXYS rats; these metabolites were not detected in the hippocampus of animals. It is important to note that during the period of pronounced neurodegenerative changes (18 months), the level of 3-methyl-2-oxovaleric acid in the serum of OXYS rats was also lower than in Wistar rats, which allows us to consider this metabolite as a potential biomarker of the early and late stages of AD. In addition, at 18 months of age, an increase in serum levels of ascorbate and a decrease in 2-ketoisovalerate, glycerol, ketoleucine, methionine, and threonine were observed in the serum of OXYS rats. In the hippocampus of 18-month-old OXYS rats, the level of threonine was, as in the serum, significantly lower than in Wistar rats, and the levels of ascorbate and glycerol did not differ (Fig. 1B).

Conclusion: The development of signs of AD in OXYS rats is preceded and accompanied by an increase in the level of scyllo-inositol and a decrease in hypotaurine in the hippocampus. Comparison of the metabolic profile of the hippocampus and rat serum revealed the presence of 34 common metabolites, and the direction of changes

with age of most of them was reciprocal in the blood and brain of the animals. At the early and late stages of development of AD-like pathology, the level of 3-methyl-2-oxovaleric acid in the blood serum of OXYS rats is reduced, which allows us to consider this metabolite as a potential available biomarker of the disease.

Funding: The study is supported by State Budget Project No. FWNR-2022-0016.

Список литературы/References

1. Kolosova N.G., Kozhevnikova O.S., Muraleva N.A., Rudnitskaya E.A., Rumyantseva Y.V., Stefanova N.A., Telegina D.V., Tyumentsev M.A., Fursova A.Z. SkQ1 as a tool for controlling accelerated senescence program: Experiments with OXYS rats. *Biochemistry (Moscow)*. 2022;87(12):1552-1562. doi 10.1134/S0006297922120124
2. Snytnikova O., Telegina D., Savina E., Tsentalovich Y., Kolosova N. Quantitative metabolomic analysis of the rat hippocampus: effects of age and of the development of Alzheimer's disease-like pathology. *J Alzheimer's Disease*. 2024;99(S2):S327-S344. doi 10.3233/JAD-230706

doi 10.18699/bgrs2024-13-20

Деконструкция механизмов старения и долголетия с помощью генных сетей

Холдина Д.А.[1]*, Дмитриев С.Е.[3], Тышковский А.Э.[2, 3]

[1] *Факультет биоинженерии и биоинформатики, Московский государственный университет, Москва, Россия*

[2] *Brigham and Women's Hospital, Медицинская Школа Гарварда, Бостон, США*

[3] *НИИ физико-химической биологии им. А.Н. Белозерского, Московский государственный университет, Москва, Россия*

** dashakholdina@gmail.com*

Ключевые слова: часы старения; транскриптомные часы; генные сети; биомаркеры старения

Мотивация и цель: В области исследования старения за последние годы наблюдаются значительные достижения, такие как разработка количественных молекулярных биомаркеров биологического возраста и продолжительности жизни [1, 2], а также открытие различных интервенций, способных увеличить продолжительность жизни млекопитающих [3, 4], как правило, путем воздействия на один или несколько установленных признаков старения [5]. Хотя молекулярные инструменты, такие как эпигенетические часы, продемонстрировали свою способность предсказывать возраст и смертность, всестороннее понимание точных механизмов, лежащих в основе этих предсказаний, остается туманным. В данной работе мы руководствовались целью деконструировать возраст-зависимые изменения генной экспрессии на функционально интерпретируемые ко-регулируемые составляющие.

Методы и алгоритмы: Чтобы создать более интерпретируемую альтернативу традиционным эпигенетическим часам, наша лаборатория разработала транскриптомные часы. Эти модели предсказывают биологический возраст и ожидаемую смертность млекопитающих, используя данные об экспрессии генов из более чем 4 500 образцов, собранных из 26 тканей мышей и крыс разного возраста и пола, а также животных, подвергшихся различным вмешательствам, регулирующим продолжительность жизни, таким как прогерия, ограничение калорийности и применение рапамицина. Для того, чтобы получить ко-регулируемые наборы генов, мы использовали данные, использовавшиеся для построения общих часов, и кластеризовали гены по характеру изменения их экспрессии с помощью метода Weighted Gene Co-expression Network Analysis (WGCNA) [6]. Далее для каждого функционально интерпретируемого модуля генов были построены собственные транскриптомные часы, которые затем применялись к данным о генной экспрессии различных возраст-зависимых заболеваний.

Результаты: Наш анализ выявил более 20 модулей генов, которые в значительной степени коррелируют со старением и смертностью, а также связаны с отдельными функциональными процессами клетки. Так, среди них были обнаружены модули, которые участвуют в воспалительной реакции, интерфероном ответе, клеточном дыхании, липидном обмене, клеточном цикле и организации внеклеточного матрикса. Используя выявленные подмножества ко-регулируемых генов, мы

построили часы хронологического возраста, смертности и продолжительности жизни для отдельных компонентов транскриптома и применили их для характеристики молекулярных механизмов, ответственных за про- или антивозрастной эффект различных возраст-зависимых заболеваний мышей (болезнь Альцгеймера, диабет и другие). В согласии с предсказанием общих часов, большинство модульных часов показало статистически значимое увеличение биологического возраста в образцах болезней по сравнению с контрольными образцами того же хронологического возраста. Модулями, внесшими наибольший вклад в повышение возраста в большинстве моделей, оказались модули воспалительного ответа, интерферонового сигнального пути и энергетического метаболизма (окислительное фосфорилирование, клеточное дыхание). Интересно, что в данных по раку, несмотря на уменьшение общего транскриптомного возраста, мы наблюдали повышение возраста для отдельных модулей, в то время как модуль, связанный с организацией внеклеточного матрикса и дифференциацией клеток, показал противоположный эффект. Известно, что хотя рак тесно связан со старением, его фенотип включает в себя как признаки старого организма, так и черты молодых де-дифференцированных клеток, которые, как было неоднократно показано ранее, ассоциированы с уменьшением биологического возраста. В контексте таких многофакторных и гетерогенных состояний, как раковые опухоли, особенно важно иметь возможность деконструировать общую оценку биологического возраста на отдельные интерпретируемые компоненты, что и отображают разработанные нами модульные часы.

Выводы: В нашей работе были идентифицированы функционально однородные ко-регулируемые модули генов, ассоциированные со старением и долголетием млекопитающих. Для этого эффективным оказалось использование метода WGCNA, благодаря которому удалось идентифицировать 26 кластеров, обогащенные генами, ответственными за воспалительный ответ, клеточное дыхание, организацию внеклеточного матрикса и другие функции. Для каждого из модулей были построены часы хронологического возраста и смертности, что позволяет деконструировать эффект композитных транскриптомных часов на отдельные функциональные компоненты и лучше интерпретировать эффект различных интервенций и заболеваний на различные механизмы старения.

From genes to gears: exploring the mechanics of aging and longevity through transcriptomic clocks and network analysis

Kholdina D.A.[1]*, Dmitriev S.E.[3], Tyshkovskiy A.E.[2, 3]

[1] *Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia*

[2] *Brigham and Women's Hospital, Harvard Medical School, Boston, USA*

[3] *Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow, Russia*

* *dashakholdina@gmail.com*

Key words: aging clocks; transcriptomic clocks; gene networks; aging biomarkers

Motivation and Aim: The field of aging research has seen significant advances in recent years, such as the development of quantitative molecular biomarkers of biological age and longevity [1, 2], and the discovery of various interventions that can increase mammalian longevity [3, 4], typically by targeting one or more established hallmarks of aging [5]. Although molecular tools such as epigenetic clocks have demonstrated their ability to predict age and mortality, a comprehensive understanding of the precise mechanisms underlying these predictions remains obscure. In this work, we were motivated by the goal of deconstructing age-dependent changes in gene expression into functionally interpretable co-regulated components.
Methods and Algorithms: To create a more interpretable alternative to traditional epigenetic clocks, our lab has developed transcriptomic clocks. These models predict mammalian biological age and expected mortality using gene expression data from more than 4,500 samples collected from 26 tissues from mice and rats of different ages and sexes, as well as animals subjected to various lifespan-regulating interventions such as progeria, caloric restriction, and rapamycin administration. In order to obtain co-regulated gene sets, we used the data used to construct a common clock and clustered genes according to the nature of their expression changes using the Weighted Gene Co-expression Network Analysis (WGCNA) method [6]. Next, a custom transcriptome clock was constructed for each functionally interpretable gene module, which was then applied to gene expression data for various age-related diseases.
Results: Our analysis identified more than 20 gene modules that are significantly correlated with aging and mortality and are also associated with distinct functional processes of the cell. Thus, among these, we found modules that are involved in inflammatory response, interferon response, cellular respiration, lipid metabolism, cell cycle and organisation of the extracellular matrix. Using the identified subsets of co-regulated genes, we constructed chronological age, mortality and lifespan clocks for individual components of the transcriptome and applied them to characterise the molecular mechanisms responsible for the pro- or anti-aging effects of various age-related diseases in mice (Alzheimer's disease, diabetes and others). In agreement with the prediction of a common clock, most of the modular clocks showed a statistically significant increase in biological age in disease samples compared to control samples of the same chronological age. The modules that contributed most to the increase in age in most models were found to be those of the inflammatory response, interferon signalling pathway and energy metabolism (oxidative phosphorylation, cellular respiration). Interestingly, in the cancer data, despite a decrease in total transcriptome age, we observed an increase in age for individual modules, while a module related to extracellular matrix organisation and cell differentiation showed the opposite effect. It is known that although cancer is closely associated with aging, its phenotype includes both features of an old organism and features of young de-differentiated cells, which have been repeatedly shown previously to be associated with a decrease in biological age. In the context of multifactorial and heterogeneous conditions such as cancer, it is particularly important to be able to deconstruct the overall estimate of biological age into individual interpretable components, which is what the modular clock we developed displays.
Conclusion: Our work identified functionally homogeneous co-regulated gene modules associated with aging and longevity in mammals. The WGCNA method was effective for this purpose, and we were able to identify 26 clusters enriched with genes responsible for inflammatory response, cellular respiration, extracellular matrix organisation and

other functions. A chronological age and mortality clock was constructed for each module, allowing us to deconstruct the effect of the composite transcriptome clock on individual functional components and to better interpret the effect of different interventions and diseases on different aging mechanisms.

Список литературы/References

1. Lu A.T., Fei Z., Haghani A. et al. Universal DNA methylation age across mammalian tissues. *Nat Aging*. 2023;3(9):1144-1166
2. Tyshkovskiy A., Ma S., Shindyapina A.V. et al. Distinct longevity mechanisms across and within species and their association with aging. *Cell*. 2023;186(13):2929-2949.e20
3. Strong R., Miller R.A., Cheng C.J. et al. Lifespan benefits for the combination of rapamycin plus acarbose and for captopril in genetically heterogeneous mice. *Aging Cell*. 2022;21(12):e13724
4. Miller R.A., Harrison D.E., Allison D.B. et al. Canagliflozin extends life span in genetically heterogeneous male but not female mice. *JCI Insight*. 2020;5(21):e140019
5. López-Otín C., Blasco M.A., Partridge L., Serrano M., Kroemer G. Hallmarks of aging: An expanding universe. *Cell*. 2023;186(2):243-278
6. Langfelder P., Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559

doi 10.18699/bgrs2024-13-23

Developing a standardized methodology to validate epigenetic aging clocks

Efimov E.[1]*†, Kriukov D.[1, 2†], Kuzmina E.[1, 2], Dylov D.V.[1, 2], Khrameeva E.E.[1]

[1] *Skolkovo Institute of Science and Technology, Moscow, Russia*

[2] *Artificial Intelligence Research Institute, Moscow, Russia*

* *Evgeniy.Efimov@skoltech.ru*

† *These authors contributed equally to this work*

Key words: aging; biomarkers; epigenetics; aging clocks; benchmarking; machine learning

Motivation and Aim: Developing a measure of human aging process becomes more and more valuable, as emerging longevity therapies require this measurement to be feasibly evaluated in clinical trials [1, 2]. However, contrary to the common machine learning models, predictors of biological age, so-called aging clocks, cannot be compared between each other in terms of their correlation with chronological age, since high accuracy of chronological age prediction does not imply high accuracy at estimating biological age. In our work, we aimed to develop a methodology to validate aging clocks by testing their ability to identify accelerated aging in a standardized panel of aging-accelerating conditions (AACs).

Methods and Algorithms: Relying on the published studies of age-related conditions, we elaborated an evidence-based set of putative AACs. To showcase our methodology, we collected a large number of open-access, age-annotated, and microarray-based datasets of human blood DNA methylation using the NCBI Gene Expression Omnibus database. We also extracted published epigenetic clock models from the respective studies and included them into our comparison pipeline. To benchmark the clock models, we proposed four tasks: two-sample aging acceleration prediction (where both healthy and AAC cohorts are available), one-sample aging acceleration prediction (in datasets containing the AAC cohort only), chronological age prediction accuracy, and systematic chronological age prediction bias. To summarize all tasks, we also suggested a cumulative benchmarking score that would penalize the results of absolute aging acceleration prediction by the ratio of prediction bias to prediction accuracy.

Results: Dataset search yielded 66 individual datasets in total comprising 19 of determined AACs. We also collected and implemented 13 existing epigenetic clock models, containing both the first generation (trained to predict chronological age only) and the second generation (trained to predict time to death) of aging clocks. The two-sample task revealed that the second-generation aging clocks (PhenoAge [3, 4] and GrimAgeV1 [5, 6] variants) appear on top, particularly at predicting aging acceleration for the immune system-related diseases (such as HIV infection and inflammatory bowel disease). Some disease classes (namely, cardiovascular and metabolic diseases) were misestimated in regards to healthy controls by all tested models. In contrast to that, first-generation aging clocks by Zhang et al. [7] and Hannum et al. [8] emerged the best at the one-sample task, in addition to the GrimAge variants [5, 6], across multiple disease classes. However, one-sample score might be compromised by systematic model error. Therefore, it was not surprising to find that, according to the cumulative benchmarking score, second-generation aging clocks still outperformed other models.

Conclusion: In our work, we developed the first systematic pipeline for validating epigenetic aging clocks based on public data to foster more rapid and robust advancements in the field of biological age research.

Funding: This work was supported by Program "Skolkovo Institute of Science and Technology – University of Sharjah Joint Projects: Artificial Intelligence for Life".

References

1. Moqri M. et al. Validation of biomarkers of aging. *Nat Med.* 2024;30(2):360-372. doi 10.1038/s41591-023-02784-9
2. Schork N.J. et al. Does Modulation of an Epigenetic Clock Define a Geroprotector? *Adv Geriatr Med Res.* 2022;4(1):e220002. doi 10.20900/agmr20220002
3. Levine M.E. et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY).* 2018;10(4):573-591. doi 10.18632/aging.101414
4. Higgins-Chen A.T. et al. A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. *Nat Aging*. 2022;2(7):644-661. doi 10.1038/s43587-022-00248-2
5. Lu A.T. et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY).* 2019;11(2):303-327. doi 10.18632/aging.101684
6. Lu A.T. et al. DNA methylation GrimAge version 2. *Aging (Albany NY).* 2022;14(23):9484-9549. doi 10.18632/aging.204434
7. Zhang Q. et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med.* 2019;11(1):54. doi 10.1186/s13073-019-0667-1
8. Hannum G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013;49(2):359-367. doi 10.1016/j.molcel.2012.10.016

doi 10.18699/bgrs2024-13-24

Polymorphic variants of autophagy regulated genes: survival rates for the aged-related pathologies and longevity

Erdman V.[1, 2]*, Petintseva A.[1, 3], Tuktarova I.[1], Timasheva Y.[1, 2], Nasibullin T.[1]

[1] *Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia*

[2] *Bashkir State Medical University, Ufa, Russia*

[3] *Ufa University of Science and Technology, Ufa, Russia*

* *danivera@mail.ru*

Key words: longevity; age-associated diseases; redox regulation; autophagy; genetic polymorphism; survival analysis

Motivation and Aim: Longevity is exceptional highly adaptive phenotype characterized by the reaching the long life expectancy while maintaining physical and cognitive abilities. Current research on longevity focuses on elucidating the endogenous and exogenous stresses manage the aging process [1]. Gradual and irreversible decline in cellular and tissue functions is observed in physiological and to a greater extent in pathological senility [2]. The achievement of longevity at the background of the age-associated diseases and the ability of long-lived individuals to adapt to a chronic pathological age background suggests the existence of adaptive endogenous mechanisms. With the better understanding of the detailed molecular mechanisms, the role of mitochondrial biogenesis and molecules that provide redox homeostasis in autophagy, which is one of the key programs for self-renewal of intracellular structures, is being strengthened [3]. The functional connection of mitochondria with cellular components involved in the process of autophagy is provided by the dynamic system of mitochondria and endoplasmic reticulum membranes [4]. The aim of our study is to search for combined polymorphic profile of genes involved in redox regulation of autophagy process, in physiological aging and age-related pathology.

Methods and Algorithms: Total sample was formed of 2806 unrelated individuals in age from 18 to 114 years and inhabiting the Republic of Bashkortostan of Russia. The mortality and survival data for the persons over 45, collected since 2001 to 2015 were built. It has allowed forming the comparison groups for assessing survival to the age of longevity and differentiating the general sample to the main clinical phenotypes based on the verified cause of death of the individual. In the groups differentiated by age and clinical status, we studied the functional polymorphic loci of genes involved in signaling pathways of enzymatic antioxidant defense, redox homeostasis, autophagy, and membrane contact components of mitochondria and endoplasmic reticulum. SNPs were genotyped by a TaqMan real-time PCR assay. Statistical analysis was carried out using the IBM SPSS Statistics (V22.0, Chicago, IL, USA), APSampler software (V.3.6.0, [5]), and R package "SNPassoc". The associations of the studied SNPs and their combinations with the age-related diseases and longevity were analyzed by the logistic regression with the different genetic models and with the Markov chain Monte Carlo method. Survival analysis was performed using the Cox proportional hazards regression.

Results: The genotype frequencies in selected SNPs of the antioxidant genes – *NFE2L2, MSRA, SOD1, SOD2, NQO1, CAT, GSR, GSTP1, GPX1,* autophagy regulated genes –

AKT1, MTOR, SIRT1, HIF1A, TP53, TOMM40, that are involved in the connection between the mitochondria and endoplasmic reticulum membrane, were established in the group of middle-aged individuals (control group, n = 1120), as well as in the groups of ageing (n = 1085) and long-lived (n = 601) persons. The obtained genotype frequencies of all studied SNPs among the control subjects were in compliance with the Hardy–Weinberg equilibrium ($p_{HWE} > 0.05$).

The old age (66–89 for men and 75–89 for women) was associated with *CAT* rs1001179*T allele (OR = 1.29 in the dominant model, p = 0.013) and *SIRT1* rs3758391*T allele (OR = 0.67 in the log-additive model, p < 0.001). Longevity (over 90) was associated with *SOD1* rs2070424*C minor allele (OR = 0.12 in the recessive model, p = 0.01). The analysis of gene-gene combinations has identified the age-related SNP's clusters. Totally, we have identified 169 multilocus patterns associated with longevity. The most significant combinations were the CAT*T+MSRA*TT+GPXI*T (OR = 5.64, $p_{FDR} = 0.002$), HIF1A*C+MSRA*T+SOD1*AA (OR = 5.2, $p_{FDR} = 0.006$), MSRA*T+SOD1*AA (OR = 4.86, $p_{FDR} = 0.012$), NQO1*C+GSR*T+SOD2*T (OR = 0.09, $p_{FDR} = 0.01$), GSTP1*A+SOD1*A+SOD2*CT (OR=3.5, $p_{FDR} = 2.93\times10^{-9}$), GSTP1*A+GSR*G (OR = 4.84, $p_{FDR} = 4.04\times10^{-9}$), and AKT1*G+HIF1A*C+GSTP1*GG (OR = 0.14, $p_{FDR} = 4.02\times10^{-9}$).

The follow-up study with the added survival analysis has revealed the elements that are core meaningful for longevity status. So, the combination of the MSRA*C and SOD1*G alleles, with the NQO1*TC genotype is rarer among long-lived women in compare with those individuals who died from all causes before reaching the longevity (OR = 0.1, $p_{FDR} = 0.035$). The cancer mortality was associated with the *CAT* rs1001179*TT (HR = 2.83, p = 0.031) and SOD1 rs2070424*AG (HR = 1.58, p = 0.018) genotype. The *PONI**G+*MSRA**C+*NQO1**T+*SOD1**G allele combination was associated with the highest risk of all-causes mortality among men (HR = 29.56, $p = 9.47\times10^{-4}$). The AKT1*GG+HIF1A*T+SIRT1*T genetic pattern had a protective effect on survival status in all studied individuals (HR = 0.71, p = 0.014).

Conclusion: We have identified the single and multilocus predictors of survival and mortality under the certain clinical phenotypes in the advanced age, as well as longevity genetic markers. We assume that these genes and their networks are involved in a stable, highly adaptive phenotype that allows to minimize the negative effects of the age-related pathologies and to reach the longevity.

Funding: The study is supported by the Russian Science Foundation grant (No. 24-25-00179).

References

1. López-Otín C., Blasco M.A., Partridge L., Serrano M., Kroemer G. Hallmarks of aging: An expanding universe. *Cell*. 2023;186(2):243-278. doi 10.1016/j.cell.2022.11.001
2. Franceschi C., Garagnani P., Morsiani C., Conte M., Santoro A., Grignolio A., Monti D., Capri M., Salvioli S. The Continuum of Aging and AgeRelated Diseases: Common Mechanisms but Different Rates. *Front Med*. 2018;5:61. doi 10.3389/fmed.2018.00061
3. Mishra J., Bhatti G.K., Sehrawat A., Singh C., Singh A., Reddy A.P., Reddy P.H., Bhatti J.S. Modulating autophagy and mitophagy as a promising therapeutic approach in neurodegenerative disorders. *Life Sci*. 2022;311(Pt. A):121153. doi 10.1016/j.lfs.2022.121153
4. Moltedo O., Remondelli P., Amodio G. The Mitochondria–Endoplasmic Reticulum Contacts and Their Critical Role in Aging and Age-Associated Diseases. *Front Cell Dev Biol*. 2019;7:172. doi 10.3389/fcell.2019.00172
5. Favorov A.V., Andreewski T.V., Sudomoina M.A. et al. A Markov chain Monte Carlo technique for identification of combinations of allelic variants underlying complex diseases in humans. *Genetics*. 2005;171(4):2113-2121. doi 10.1534/genetics.105.048090

doi 10.18699/bgrs2024-13-25

Research of non-monotonic aging changes based on multiomics data

Gorbarenko A.V.[1], Bobkov G.A.[1, 2], Alekseev A.A.[3, 4, 5]*

[1] *Bioinformatics Institute, St. Petersburg, Russia*

[2] *ITMO, St. Petersburg, Russia*

[3] *Lomonosov Moscow State University, Moscow, Russia*

[4] *Kulakov National Medicine Center for Obstetrics, Gynecology and Perinatology, Moscow, Russia*

[5] *Russian Research Institute of Health, Moscow, Russia*

* *alekseev@physics.msu.ru*

Key words: aging; aging-related trends; omics data; aging transitions; trends clustering

Motivation and Aim: At the age around 37, various molecular alterations occur in human organisms, including dysregulated gene expression, altered metabolite levels, somatic mutations, epimutations, and accumulated molecular damage [7-10]. Age transition in 23±2 years old (y.o.) is also very important for the system biology of aging: first increasing of MDA (malondialdehyde, marker of oxidative stress) occurs at this age [12]. Significant stages in the aging process also occur around ages 24 and 60, where molecular and physiological changes become most pronounced [11-13]. Wide known "hallmarks of aging" involve genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulation of nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication. These mechanisms contribute to aging at molecular, cellular, and systemic levels [6, 14, 15], but interaction between cellular and systemic aging is still unclear. This study aims to obtain the most relevant cellular aspect, associated with metabolic transition of aging around 37 through multiomics analysis, including epigenomics and RNA-sequencing data.

Methods and Algorithms: Data was collected from the National Center for Biotechnology Information Gene Expression Omnibus database, accessed via the geoparse library [1]. The complete datasets table can be found on GitHub. All CpG data were selected based on the platform, specifically only using the Illumina 450k array. Due to the large size of the methylation data, only CpGs that significantly differ between the age groups '18-27', '27-30', '30-35', '35-40', '40-45', or '45-60' years were chosen for further investigation. Batch effects were detected using PCA visualization and corrected with pyCombat [2]. The most representative dataset GSE42861was chosen to test methods for calculating central tendency (mean, median, density distribution peak) and clustering algorithms (k-means, hierarchical, DBSCAN, Gaussian Mixture). Correlations of CpG trends were calculated using the Pearson correlation coefficient. The chosen algorithm was applied to all methylation and microarray groups of datasets. Additionally, separate analyses were performed for males and females. The clustering results were visualized using a sliding window to plot samples from each cluster. CpG annotation was performed using Illumina's annotation data [3]. Data from epigenetics and transcriptomics were intersected. Enrichment analysis was conducted using Enrichr (Reactome database) and StringDB tools to identify pathways with the most significant changes [4, 5].

Results: Based on the GSE42861 dataset, both the median and density distribution peak methods yielded similar results, but the density distribution peak showed better stability within the methylation group of datasets. Hierarchical clustering and GM produced comparable results, with GM demonstrating better stability. Therefore, the density distribution peak was chosen for central tendency, and GM was selected as the clustering algorithm. The GM covariance matrix was calculated as diagonal due to computational constraints. We chose eight clusters to capture additional points of age-related changes (Fig. 2), exceeding the optimal five. The clusters with changes, obtained as described above, identified 3354 CpGs and 290 RNAs around the age of 24. As a result, we identified 26 genes. The full list of genes is available on GitHub. Using StringDB (Fig. 1), we observed that these genes are responsible for changes in cellular regulation, autophagy, cell cycles, and more. Enrichr analysis showed that the top two enrichments were in FLT3 signaling and cellular senescence. For the male datasets, the shape and age distribution were insufficient to conduct a robust analysis of the methylation data. A separate analysis of the transcriptomic data revealed transitions at ages 37-40 in both males and females, specifically involving the genes PERP, CDKN1A, THBS1, TFPI, and SERPINB.

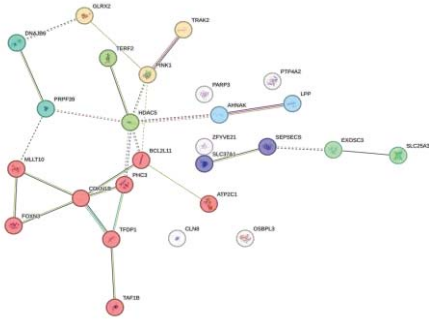


Fig. 1. Genes clustering using StringDB

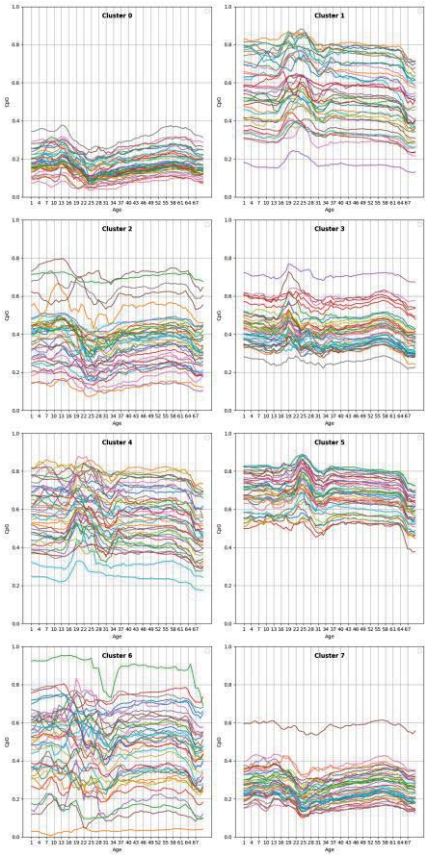


Fig. 2. Methylation trends clustering

Conclusion: Our study aimed to uncover key molecular pathways and biological processes that undergo significant changes at around 24 and 37 y.o. However, it is important to note that we were unable to identify significant changes at 37 y.o. based on epigenetic analysis alone. This suggests that while transcriptomic data shows clear transitions, epigenetic markers might not be as sensitive or may require different analytical approaches to detect changes at 37 y.o. Identified 26 key genes play distinct roles in cellular processes. CDKN1B regulates cell cycle progression and contributes to cellular senescence. FOXN3 acts as a transcriptional repressor, influencing DNA damage pathways. BCL2L11 induces apoptosis and is crucial for removing damaged cells. HDAC5 regulates gene expression, impacting tissue function and development. PHC3 maintains transcriptional repression, influencing differentiation and senescence. TAF1B is involved in transcription initiation, affecting aging-related processes. Other genes contribute to functions like pre-mRNA splicing and mitochondrial activity, collectively

influencing aging through diverse mechanisms. For future investigations, collecting more datasets, including lipidomics and metabolomics data, is essential for a comprehensive understanding of aging's molecular changes. In the next steps of our research, reviewing our analytical pipeline in transcriptomic data and improving correlation calculations for all data could also enhance our analysis. But even at this stage our study contributes to understanding aging by identifying key genes and pathways undergoing significant changes around 24 y.o.

Data availability: A repository was created on GitHub, containing a detailed progress of work and the results obtained. The repository can be found at: https://github.com/CaptnClementine/SODA.

References

1. Gene Expression Omnibus (GEO) database, National Center for Biotechnology Information (NCBI). Available at: https://www.ncbi.nlm.nih.gov/geo/
2. Johnson W.E., Li C., Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118-127
3. Illumina, Inc. (n.d.). Infinium HumanMethylation450 BeadChip Annotation. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_humanmethylation450.pdf
4. Kuleshov M.V., Jones M.R., Rouillard A.D. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90-W97
5. Szklarczyk D., Gable A.L., Lyon D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-D613
6. Mc Auley M.T., Guimera A.M., Hodgson D. et al. Modelling the molecular mechanisms of aging. *Biosci Rep.* 2017;37(1):BSR20160177
7. Aramillo Irizar P., Schäuble S., Esser D. et al. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat Commun.* 2018;9(1):327
8. Kabacik S., Lowe D., Fransen L. et al. The relationship between epigenetic age and the hallmarks of aging in human cells. *Nat Aging.* 2022;2(6):484-493
9. Horvath S., Gurven M., Levine M.E. et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol.* 2016;17:1-23
10. González-Velasco O., Papy-García D., Le Douaron G. et al. Transcriptomic landscape, gene signatures and regulatory profile of aging in the human brain. *Biochim Biophys Acta.* 2020;1863(6):194491
11. Zavorsky G.S., Tomko K.A., Smoliga J.M. Declines in marathon performance: Sex differences in elite and recreational athletes. *PloS One.* 2017;12(2):e0172121
12. Ravera S., Podestà M., Sabatini F. et al. Discrete changes in glucose metabolism define aging. *Sci Rep.* 2019;9(1):10347
13. Lu A.T., Binder A.M., Zhang J. et al. DNA methylation GrimAge version 2. *Aging (Albany NY).* 2022;14(23):9484
14. Yaneske E., Angione C. The poly-omics of ageing through individual-based metabolic modelling. *BMC bioinformatics*. 2018;19:83-96
15. Lehallier B., Gate D., Schaum N. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat med.* 2019;25(12):1843-1850

doi 10.18699/bgrs2024-13-26

Epigenetics and accelerated aging of Far Northern Yakutian populaiton

Kalyakulina A.[1, 2], Yusipov I.[1, 2], Kondakova E.[1, 3], Ivanchenko M.[1, 2]*

[1] *Laboratory of Systems Medicine of Healthy Aging, Institute of Biogerontology, Lobachevsky University, Nizhny Novgorod, Russia*

[2] *Department of Applied Mathematics, Institute of Information Technologies, Mathematics and Mechanics, Lobachevsky University, Nizhny Novgorod, Russia*

[3] *Institute of Neuroscience, Lobachevsky University, Nizhny Novgorod, Russia*

* *ivanchenko@unn.ru*

Key words: aging; epigenetics; adaptation

We present the results of the first epigenetic study of the Yakutian population using whole-blood DNA methylation data, supplemented with the comparison to the residents of Central Russia [1]. In particular, we aimed to improve understanding of the mechanism of accelerated aging, development of age-related diseases in extreme climate conditions, as well as adaptation of organism to the cold. Gene set enrichment analysis revealed, among others, geographic region-specific differentially methylated regions associated with adaptation to climatic conditions (water consumption, digestive system regulation), aging processes (actin filament activity, cell fate), and both of them (channel activity, regulation of steroid and corticosteroid hormone secretion). Further, it is demonstrated that the epigenetic age acceleration of the Yakutian representatives is significantly higher than that of Central Russia counterparts. For both geographic regions, we showed that epigenetically males age faster than females, whereas no significant sex differences were found between the regions.

These findings promote the fundamental understanding of the mechanisms of modulating gene activity and functional clustering of the involved gene loci associated with the age-related changes in metabolism of the residents of various regions of the Russian Federation. It also paves the way to assessing the risk of developing age-associated diseases in people living in extreme climatic conditions, and provides the basis for the development of preventive approaches and age control schemes.

References

1. Kalyakulina A., Yusipov I., Kondakova E., Bacalini M.-G., Giuliani C., Sivtseva T., Semenov S., Ksenofontov A., Nikolaeva M., Khusnutdinova E., Zakharova R., Vedunova M., Franceschi C., Ivanchenko M. Epigenetic features of far northern Yakutian population. *Clin Epigene*. 2023;15:189. doi 10.1186/s13148-023-01600-y

doi 10.18699/bgrs2024-13-27

Antioxidants: good, bad, or ineffective in terms of aging and lifespan? It depends

Khokhlov A.N.*

Evolutionary Cytogerontology Sector, School of Biology, Lomonosov Moscow State University, Moscow, Russia

** khokhlovan@my.msu.ru*

Key words: aging; lifespan; free radicals; nutrition; experimental gerontology; synthetic and natural antioxidants; age-related diseases

The idea that antioxidants (AO) can be effective geroprotectors became widespread in the second half of the 20th century. It was based on the concept of the leading role of free radicals (FR) – mainly reactive oxygen species (ROS) – in the aging process, formulated by Denham Harman and Nikolay Emanuel. Subsequently, a huge number of experimental works appeared, which substantiated the point of view according to which the introduction into the body of compounds that scavenge FR leads to a slowdown in aging and an increase in lifespan. Actually, for many years, gerontologists used the terms geroprotectors/anti-aging drugs to mainly designate AO of various kinds. In this case, as a rule, we were talking about synthetic AO, the doses of which could be quite easily controlled.

It should be noted that the term "geroprotectors" was coined by scientists from the Institute of Chemical Physics of the USSR Academy of Sciences in the last quarter of the former century specifically in relation to the synthetic AO created in this institution. We studied one of them, epygid (2-ethyl-6-methyl-3-hydroxypyridine hydrochloride), quite intensively in terms of its effect on stationary phase aging of cultured cells of different nature [1, 2]. It turned out that epygid acts differently on different cells. In particular, we were unable to use it to slow down the stationary phase aging of normal embryonic diploid human fibroblasts. As a result, we already then concluded that it is impossible to help "young" cells with high viability and proliferative activity to age more slowly using AO. Unless, of course, you specifically stimulate in them oxidative stress in one way or another, which is used quite often at the present stage of gerontological research [3, 4].

Numerous studies of the geroprotective properties of AO in experiments on laboratory animals, at first glance, have demonstrated their significant effectiveness in slowing down aging and increasing lifespan, both average and maximum. However, firstly, it remained unclear how applicable the data obtained could be for use in humans – including for suppressing the development of age-related diseases such as sporadic Alzheimer's disease [5] – and secondly, many questions arose regarding the correct selection of control objects in such experiments [6]. As a rule, these were animals that either had certain genetic abnormalities or were in unfavorable living conditions, which led to the development of oxidative stress in them. Obviously, AO were quite effective in such cases. The interpretation of the data obtained was also influenced by the still existing differences in the understanding of what aging is [7]. Very often, the classical definition of aging as a set of age-related changes in the body, leading to an increase in the probability of its death, is simply ignored.

We must not forget that FR in the body are often simply necessary for its normal functioning, so it cannot be said that their elimination is always beneficial [8, 9]. In particular, an entire special issue of the journal "Oxidative Medicine and Cellular Longevity" entitled "Harmful and Beneficial Role of ROS 2020" was devoted to this problem [10].

It is also necessary to emphasize that if an overdose of natural AO entering our body with food is practically impossible, then in the case of synthetic AO [11] it becomes quite real and occurs in very many cases, since it is very difficult to determine their concentration in organs and tissues when taking medications with water or food. This, on the one hand, can lead, paradoxically, to the development of oxidative stress (the so-called "antioxidant paradox") [12], and on the other, suppress the corresponding functions for which FR are necessary [9]. In particular, it was shown that increased mitochondrial ROS production plays an integral role in the acute β-adrenergic-induced inotropic response of cardiomyocytes. This stimulatory effect is mediated via cAMP–PKA-dependent and Ca2+-independent signaling [13]. It is also known that FR (both ROS and reactive nitrogen species) are needed for the normal functioning of the immune system, which provides protection against various infections, as well as for the processes of maturation of cellular structures and for maintaining the functioning of cellular signaling systems [14]. In particular, it is known that phagocytes (neutrophils, macrophages, monocytes) release free radicals to destroy invading pathogenic microbes as part of the body's defense mechanism against disease [15].

It should be noted that there is even a paradoxical, at first glance, point of view, according to which pro-oxidants in some cases may be more beneficial for the body than AO [16].

Thus, it seems that the popular idea of AO as a panacea against "normal" aging today cannot be considered absolutely proven. Apparently, with the help of AO it is possible to quite successfully combat oxidative stress that occurs in various pathologies, and thereby increase the average lifespan (sometimes even the maximum one). It can be assumed that in some cases this occurs due to a reduction in the "oxidative-stress" manifestations of age-related diseases, which is also not bad, although, apparently, it does not bring us closer to realizing the dream of a "fountain of eternal youth."

Conclusion: So maybe we should still adopt – at least in some cases – the slogan "Freedom for the radicals!".

Funding: This work was supported by ongoing institutional funding. No additional grants to carry out or direct this particular research were obtained.

References

1. Khokhlov A.N., Ushakov V.L., Kapitanov A.B., Nadjariyan T.L. Effect of the 2-ethyl-6-methyl-3-hydroxypyridine hydrochloride geroprotector on proliferation of Acholeplasma laidlawii cells. *Doklady AN SSSR*. 1984;274(4):930-933 (in Russian)
2. Khokhlov A.N., Golovina M.E., Chirkova E.Yu., Nadjariyan T.L. Analysis of some kinetic characteristics of cultured cell growth. III. Effects of inoculation density, geroprotector-antioxidant, "stationary phase ageing". *Tsitologiya*. 1987;29(3):353-357 (in Russian)
3. von Zglinicki T., Saretzki G., Döcke W., Lotze C. Mild hyperoxia shortens telomeres and inhibits proliferation of fibroblasts: a model for senescence? *Exp Cell Res*. 1995;220(1):186-193. doi 10.1006/excr.1995.1305
4. Saretzki G., Feng J., von Zglinicki T., Villeponteau B. Similar gene expression pattern in senescent and hyperoxic-treated fibroblasts. *J Gerontol Series A Biol Sci Med Sci.*1998;53(6):B438-B442. doi 10.1093/gerona/53A.6.B438
5. Khokhlov A. N. Why freshwater hydra does not get Alzheimer's disease. *Moscow University Biological Sciences Bulletin*. 2023;78(3):198-204. doi 10.3103/S0096392523700104

6. Khokhlov A.N., Klebanov A.A., Morgunova G.V. On choosing control objects in experimental gerontological research. *Moscow University Biological Sciences Bulletin*. 2018;73(2):59-62. doi 10.3103/S0096392518020049
7. Khokhlov A.N. Gerontology in the 21st century: From failures to advances. Hopefully. *Adv Gerontol*. 2023;13 (1):1-3. doi 10.1134/S2079057024600289
8. McCann S. Antioxidants: good, bad or indifferent. *Bone Marrow Transplantation*. 2019;54(1):1-2. doi 10.1038/s41409-018-0397-9
9. Di Meo S., Venditti P. Evolution of the knowledge of free radicals and other oxidants. *Oxid Med Cell Longevity*. 2020;2020:9829176. doi 10.1155/2020/9829176
10. Di Meo S., Venditti P., Victor V.M., Napolitano G. Harmful and beneficial role of ROS 2020. *Oxid Med Cell Longevity*. 2022;2022:9873652. doi 10.1155/2022/9873652
11. Stokes P., Belay R.E., Ko E.Y. Synthetic antioxidants. In: Male Infertility: Contemporary Clinical Approaches, Andrology, ART and Antioxidants. Springer, 2020;543-551. doi 10.1007/978-3-030-32300-4_44
12. Henkel R., Agarwal A. Harmful effects of antioxidant therapy. In: Male Infertility: Contemporary Clinical Approaches, Andrology, ART and Antioxidants. Springer, 2020;845-854. doi 10.1007/978-3-030-32300-4_68
13. Andersson D.C., Fauconnier J., Yamada T., Lacampagne A., Zhang S.J., Katz A., Westerblad H. Mitochondrial production of reactive oxygen species contributes to the β-adrenergic stimulation of mouse cardiomycytes. *J Physiol*. 2011;589(7):1791-1801. doi 10.1113/jphysiol.2010.202838
14. Pham-Huy L.A., He H., Pham-Huy C. Free radicals, antioxidants in disease and health. *Int J Biomed Sci*. 2008;4(2):89-96
15. Dröge W. Free radicals in the physiological control of cell function. *Physiol Rev*. 2002;82(1):47-95. doi 10.1152/physrev.00018.2001
16. Carocho M., Ferreira I.C. A review on antioxidants, prooxidants and related controversy: Natural and synthetic compounds, screening and analysis methodologies and future perspectives. *Food Chemical Toxicol*. 2013;51:15-25. doi 10.1016/j.fct.2012.09.021

doi 10.18699/bgrs2024-13-28

Free radicals, antioxidants and longevity: aging versus reliability of biological systems

Koltover V.
Federal Research Center of Problems of Chemical Physics and Medical Chemistry, RAS, Chernogolovka, Moscow Region, Russia
koltover@icp.ac.ru

Key words: aging; longevity; free radicals; antioxidants; system biology; reliability; robustness

Motivation and Aim: Aging is a universal process to which all organisms, multi-cellular and unicellular, are subjected. Some researchers consider aging as the last stage of a genetic program of ontogenesis and believe that there are special "genes of aging" which regulate aging and death. Other researchers believe that aging is a stochastic process, the cumulative side result of the pleiotropic action of a large number of regulatory genes, whose useful action manifests itself in the reproductive period and whose harmful influence becomes apparent only after the reproductive period ends. This mini-review is designed to show that the system theory of reliability allows to integrate the programmed events and the stochastic events into the single united theory.
Methods and Algorithms: In engineering, reliability is defined as ability of a device to perform the preset function for the given time under the given conditions [1, 2]. Smilitarily to technical devices, biological constructs are not perfectly reliable in operation, i. e. normal acts of operations alternate with stochastic (random) malfunctions (failures). The field of systems biology, in dealing with the problem of reliability, incorporates investigations of: classification and systematization of failures; mechanisms of failures of biomolecular nanoreactors and mechanisms of realization of the failures in functional breaks; investigations of renewal processes; elaboration of methods for testing reliability and predicting failures. This field was born about 50 years ago. Regular conferences on reliability of biological systems, starting from 1975 in the former USSR, have given the impetus to research in this direction [3]. The special Committee on Reliability of Biological Systems at the Scientific Council on Biological Physics of the USSR Academy of Sciences, to deal with the problems of reliability of biological systems, was organized in 1978, and many prominent biophysicists were the members of this Committee. A quarter of a century after, it has spurred the similar studies behind the former iron curtain under the style of "biological robustness" [4]. Yet, there is the mathematical theory of reliability, not "robustness", and mathematicians consider this theory as a part of probability theory [1, 2]. As said by Immanuel Kant, "in any partial doctrine of nature, one can find as much genuine science as there is mathematics in this doctrine and no more". The problem of reliability has direct bond to the problem of aging of biological systems. The systems reliability approach, which was developed in our papers, is based on several general postulates [5–9]. First, all biological constructs are designed in keeping with the genetic program in order to perform the preset functions. Second, we believe that all constructs operate with the limited reliability, namely, for each and every biological device normal operation acts alternate with accidental malfunctions (recurrent failures). Third, the preventive maintenance, i. e., the timely replacement (prophylaxis) of unreliable functional elements through the

metabolic turnover, that follows the pattern preset in the genome, is the main line of assuring the high systems reliability. Forth, there are a finite number of critical elements of the highest hierarchic level which perform the supervisory functions over the preventive maintenance. And, five, the supervisors also operate with the genetically preset and limited reliability.

Results: On the reliability-engineering basis the universal features of aging, the exponential growth of mortality rate with time (Gompertz law of mortality) and the correlation of longevity with the species-specific resting metabolism (Rubner scaling relation) are explained. The stochastic malfunctions of the mitochondrial electron transport nanoreactors, which produce the anion-radicals of oxygen ("superoxide radicals") as by-products of oxidative phosphorylation, are of first importance. As the reducing agent, this radical affects NAD(P)H/NAD(P)$^+$ ratio, thereby impacting the epigenetic sirtuin regulators of the metabolic repair and renewal processes. As the consequence, the oxidative-stress products and other metabolic slags accumulate with time, resulting the impetus to autophagic or apoptotic cell death with age-associated clinical disorders. Longevity of human brain could reach 250 years should the antioxidant enzyme defence against the free-radical failures be perfect. Furthermore, the systems reliability approach serves as the heuristic methodology for anti-aging pharmacology, in part, in searching the real protection mechanisms of the antioxidant therapy [10]. *In vivo*, the so-called antioxidants, natural and synthetic ones, work not so much as the direct inhibitors of the free radical processes but prevent the formation of the oxygen radicals and free radical oxidation processes in cells and tissues. In terms of the system theory of reliability, antioxidants provide the prophylactic maintenance against reactive oxygen species. For example, synthetic antioxidant BHT (butylated hydroxytolene) prevents generation of the oxygen radicals as by-products of mitochondrial electron transport, while resveratrol and some other flavonoids exert the preventive antioxidant action by inducing the expression of the antioxidant enzymes. At this point, the neuro-hormonal system and regulatory transcriptional and translation factors, along with the organism's microbiota, are of paramount importance.

Conclusions: Thus, from the reliability point of view, aging occurs as the inevitable consequence of the genetically preset deficiency in reliability of the biomolecular constructs while the free-radical redox-timer, located presumably in the special cells of hypothalamus, serves as the effective stochastic mechanism of realization of the genetics program. Furthermore, the systems reliability approach serves as the heuristic methodology for searching the real mechanisms of anti-aging pharmacology, including the antioxidant therapy.

Funding: The study is supported by the Ministry of Science and Higher Education of Russian Federation (theme FFSG-2024-0012).

References

1. Lloyd D.K., Lipov M. Reliability: Management, Methods and Mathematics. New Jersey: Prentice Hall Inc, 1962
2. Gnedenko B.V., Belyaev Yu.K., Soloviev A.D. Mathematical Methods in Theory of Reliability. Moscow: Nauka, 1965 (in Russian)
3. Grodzinsky D.M., Vojtenko V.P., Kutlakhmedov Yu.A., Koltover V.K. Reliability and Aging of Biological Systems. Kiev: Naukova Dumka, 1987 (in Russian)
4. Kitano H. Biological robustness. *Nat Rev Genet*. 2004;5:826-837
5. Koltover V.K. Reliability of enzymatic protection of a cell against superoxide radicals and the aging. *Doklady Biophysics (Doklady Akad Nauk SSSR)*. 1981;256(1):3-5
6. Koltover V.K. Free radical theory of aging: View against the reliability theory. In: Free Radicals and Aging. Basel: Birkhauser, 1992;11-19

7. Koltover V.K. Reliability concept as a trend in biophysics of aging. *J Theor Biol.* 1997;184(2):157-163
8. Koltover V.K. Mathematical theory of reliability and biological robustness: Reliable systems from unreliable elements. In: Focus on Systems Theory Research. New York: Nova Science Publishers Inc., 2019;49-80
9. Koltover V.K. Reliability and longevity of biological systems: the free-radical redox timer of aging. In: Redox Signaling and Biomarkers in Ageing. Healthy Ageing and Longevity. Cham: Springer, 2022;21-44
10. Koltover V.K., Skipa T.A. Antioxidant Pharmacology. In: Anti-Aging Pharmacology. Elsevier Acad. Press (USA), 2023;341-366

doi 10.18699/bgrs2024-13-29

Epistemic uncertainty challenges reliable prediction of rejuvenation events with aging clocks

Kriukov D.O.[1]*, Kuzmina E.A.[1], Efimov E.O.[1], Dylov D.V.[1, 2], Khrameeva E.E.[1]

[1] *Skolkovo Institute of Science and Technology, Moscow, Russia*

[2] *Artificial Intelligence Research Institute, Moscow, Russia*

* *dmitrii.kriukov@skoltech.ru*

Key words: rejuvenation; cell reprogramming; epistemic uncertainty; epigenetic aging clocks; dataset shift; DNA methylation

Motivation and Aim: Epigenetic aging clocks have been widely used to validate rejuvenation events during cellular reprogramming [1] and early stages of embryogenesis [2]. However, the predictions of aging clocks can not be directly verified due to the unavailability of true biological ages for reprogramming or embryonic methylation samples. Moreover, the extreme nature of reprogramming and embryogenesis data, compared to the data used for aging clock training, raises questions about the reliability of aging clock predictions (Fig. 1a, b).

Methods and Algorithms: We turned to established knowledge in machine learning and constructed a multifaceted analytical framework to consider rejuvenation predictions from the uncertainty perspective [3] (Fig. 1c–g).

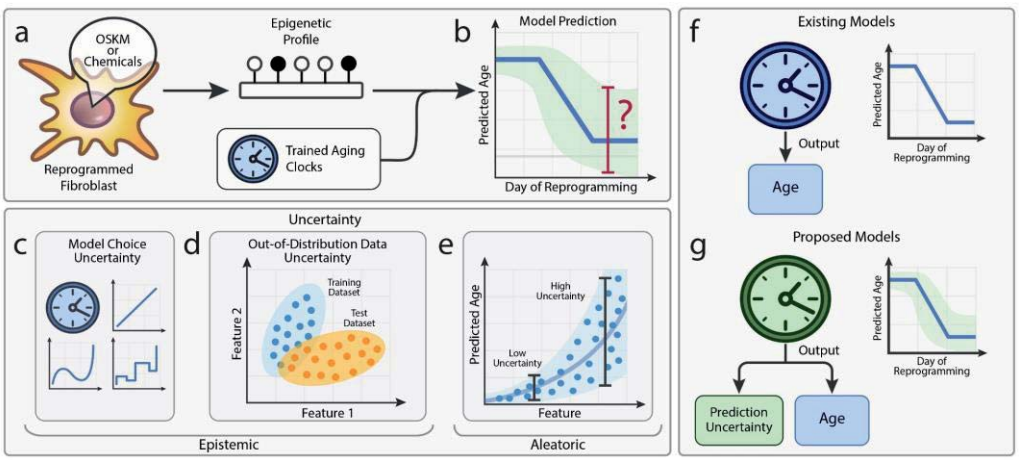


Fig. 1. Prediction uncertainty is an essential component for clinically relevant aging clocks. a, A common pipeline for testing rejuvenation effect using epigenetic aging clocks. b, Current aging clocks estimate biological age during the reprogramming process; however, they lack epistemic uncertainty quantification. c, By selecting a model to make predictions from data, a researcher implicitly introduces model uncertainty. d, Out-of-distribution uncertainty arises when the testing data samples are not represented in the training distribution. e, Aleatoric uncertainty comes from the intrinsic variability in data, e. g. when the same level of a feature corresponds to different ages. f, None of the existing aging clocks estimates epistemic uncertainty. g, We propose to use aging clocks capable of predicting uncertainty, which could mitigate the potentially erroneous effects of clock predictions on clinical decision-making

The framework includes: testing two datasets (training and testing) on the distribution shift [4], testing consistency of rejuvenation prediction by training multiple machine learning models from different model families, direct uncertainty estimation with Gaussian Process Regression model [5]. This framework addresses four key questions: (i) Is there a substantial shift in DNAm values between aging and rejuvenation datasets? (ii) Do different established aging clocks provide consistent predictions for rejuvenation? (iii) Can the rejuvenation datasets be used reciprocally to predict normal aging? (iv) Can an aging clock, capable of estimating its own uncertainty, demonstrate age reversal in putative rejuvenation experiments?

Results: Our analysis shows that DNA methylation profiles of reprogramming are not represented in the aging data used for clock training – a phenomenon also known as dataset shift – which introduces high epistemic uncertainty in aging predictions. Furthermore, predictions of different published aging clocks exhibit poor agreement, with some even suggesting zero or negative rejuvenation. Our developed Inverse Train-Test Procedure further reveals that in vitro reprogramming datasets can not robustly predict normal aging reciprocally, thus challenging the assumption that normal aging can precisely predict reprogramming. We demonstrate that the high prediction uncertainty challenges the reliability of aging clocks in predicting rejuvenation effects observed during in vitro reprogramming prior to pluripotency and throughout embryogenesis. Conversely, our method also reveals a significant age increase after in vivo reprogramming.

References

1. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115
2. Trapp A., Kerepesi C., Gladyshev V.N. Profiling epigenetic age in single cells. *Nat Aging*. 2021;1(12):1189-1201
3. Chua M. et al. Tackling prediction uncertainty in machine learning for healthcare. *Nat Biomed Eng*. 2023;7(6):711-718
4. Dataset Shift in Machine Learning. MIT Press, 2008
5. Williams C., Rasmussen C. Gaussian Processes for Regression. In: Touretzky D.S., Mozer M.C., Hasselmo M.E. (Eds.). Advances in Neural Information Processing Systems 8. MIT Press, 1996;514-520

doi 10.18699/bgrs2024-13-30

Knockout of *Hsp70* genes causes changes in the transcriptome of leg skeletal muscles and locomotion speed in *D. melanogaster* comparable to age-related changes

Kurochkina N.S.[1]*, Makhnovskii P.A.[1], Kukushkina I.V.[2], Popov D.V.[1]

[1] *Institute of Biomedical Problems, RAS, Moscow, Russia*

[2] *Moscow State University, Moscow, Russia*

* *nadia_sk@mail.ru*

Key words: skeletal muscle; aging; heat shock proteins; transcriptome

Motivation and Aim: Aging is associated with complex changes, including a decrease in physical activity, mass and functional capabilities (insulin sensitivity, performance and muscle strength) of skeletal muscles. These changes are based on various molecular and cellular processes, among which the disturbance of proteostasis plays an important role. Heat shock proteins 70 (HSP70) play a key role in the regulation of proteolysis. The association of *Hsp70* (*Hspa*) genes expression in skeletal muscle with primary aging has recently been discovered [1], but their role in age-related changes in skeletal muscle functions is not clear [2]. The aim of the work was to study the effect of knockout of *Hsp70* genes on transcriptome of leg skeletal muscle and their performance and compare these changes with age–related changes in the control group (without knockout).
Methods and Algorithms: Knockout is a classic approach for studying the function of a gene. The *Hsp70* family consists of more than 10 genes; knockout of one of them can be compensated by other genes of this family, which makes it difficult to interpret the results of such experiments. Therefore, we used the *D. melanogaster* line with knockout of 6 of the 13 genes in this family. The study was performed on males of *D. melanogaster* control line w^{1118} and line $Hsp70^{-}$ with knockout of *Hsp70Aa, Hsp70Ab, Hsp70Ba, Hsp70Bb, HSP70BB, Hsp70Bc* genes (orthologs of the *Hspa1a, Hspa1b, Hspa2* and *Hspa8* mammalian genes). Limbs consisting mainly of skeletal muscles were separated from 7-, 23- and 47-day-old flies of both lines, then RNA was isolated to assess the transcriptomic profile of protein-coding genes by RNA sequencing in single-end mode with a depth of 25 million reads/sample (NextSeq 550, Illumina, USA). Differentially expressed genes were defined as genes with an expression level >1 TPM, *Padj* <0.01 and fold change >1.25 times. The performance of flies was assessed by their locomotion speed (climbing a tube wall) in a negative geotaxis test, continuously determining the position and vertical speed of each fly (FreeClimber tool).
Results: In old w^{1118} flies (47 days vs. 7 days), the locomotion speed decreased significantly and the leg transcriptomic profile changed (more than 3,500 genes). Functional enrichment analysis showed that up-regulated genes belonged to functional categories related to nuclear proteins, proteolysis/proteasomes, protein folding, translation/ribosomes and extracellular matrix. Down-regulated genes were associated with mitochondrial proteins (oxidative enzymes, etc.), membrane proteins and synaptic proteins of motor neurons, enzymes of glucose/glycogen and amino acid metabolism, proteins of sarcomeres and extracellular matrix. These changes are in good agreement with age-related changes in the transcriptome of skeletal muscles in rodents and humans [3, 4].

Knockout of six genes of the *Hsp70* family led to a decrease in the locomotion speed in young (7-day-old) flies by 2 times, relative to young control flies (w^{1118} line), which was accompanied by noticeable changes in the transcriptomic profile of leg muscles (more than 1,500 genes). Interestingly, about a third of these genes also changed expression during aging in the control group. Among them, up-regulated genes enriched the functional category "extracellular space", and down-regulated genes – "plasma membrane", "metabolic pathways", "carbohydrate metabolism".

Conclusion: Knockout of six genes of the *Hsp70* family causes pronounced changes in the transcriptome of leg skeletal muscles and performance (locomotion speed) in young (7-day-old) flies. A third of these transcriptomic changes overlapped with age-related changes in the control line (47 days vs. 7 days), which partially explains the decrease in locomotion speed in young $Hsp70^{-}$ flies relatively young w^{1118} flies.

Funding: The study was supported by the Russian Science Foundation grant No. 21-15-00405.

References

1. Kurochkina N.S., Orlova M.A., Vigovskiy M.A. et al. Age-related changes in human skeletal muscle transcriptome and proteome are more affected by chronic inflammation and physical inactivity than primary aging. *Aging Cell*. 2024;23(4):e14098
2. Goh J., Wong E., Soh J. et al. Targeting the molecular & cellular pillars of human aging with exercise. *FEBS J*. 2023;290(3):649
3. Borsch A., Ham D.J., Mittal N. et al. Molecular and phenotypic analysis of rodent models reveals conserved and species-specific modulators of human sarcopenia. *Commun Biol*. 2021;4(1):194
4. Su J., Ekman C., Oskolkov N. et al. A novel atlas of gene expression in human skeletal muscle reveals molecular changes associated with aging. *Skelet Muscle*. 2015;5:35

doi 10.18699/bgrs2024-13-31

Development a simple computational pipeline for processing Illumina methylation data

Lopatenko M.V.
Institute of Biology and Biomedicine, Lobachevsky State University of Nizhni Novgorod, Nizhni Novgorod, Russia
fasdan1999@gmail.com

Key words: bioinformatics pipeline; epigenetics

In the context of continuous growth of DNA methylation data due to requests in gerontology and oncology, there is a need to improve the methods of their processing. At the same time, there is a demand for a simple and convenient solution that aggregates different software tools to increase the variability and accuracy of the analytical process [1]. Thus, we present a DNA methylation data processing pipeline that does not require a high threshold of entry, making it accessible to a wide range of scientific community.
In this work, the developed pipeline was applied on the IDAT file (Illumina arrays raw data) from Illumina's Infinium arrays (27k, 450k, and EPIC), obtained by accessing the Zenodo database (https://zenodo.org/).
The pipeline itself included algorithms and libraries in R and Python for error correction, data normalization and identification of sites with differential methylation and had the following steps: 1) preparation of methylation IDAT files; 2) quality control (QC) and filtering analysis; 3) analysis of processed data and interpretation.
The obtained result can facilitate the construction of age calculation models for epigenetic clocks by integrating the developed pipeline with machine learning algorithms. Moreover, the pipeline is very flexible and can be modified to work with bisulfite sequencing (BS-seq) data, including their quality control, alignment, methylation calling, analysis and visualization of results. As a result, it will significantly increase the functionality and quality of the processed information.

References

1. Dafni A., Piferrer F. Bioinformatic analysis for age prediction using epigenetic clocks: Application to fisheries management and conservation biology. *Front Marine Sci.* 2023;10:1096909

doi 10.18699/bgrs2024-13-32

Epigenetic age acceleration and colorectal cancer in an ageing population cohort

Malyutina S.[1]*, Chervova O.[2], Maximov V.[1], Nikitenko T.[1]

[1] *Scientific Research Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics, SB RAS (IIPM – Branch IC&G, SB RAS), Novosibirsk, Russia*

[2] *University College London, London, United Kingdom*

* *smalyutina@hotmail.com*

Key words: DNA methylation; epigenetic age; colorectal cancer; ageing; population

Motivation and Aim: Ageing is one of the most important risk factor of colorectal cancer (CRC) [1], and its contribution largely depends on 'biological age' which reflects an individual rate of health decline. The measure of 'epigenetic age' (EA) based on DNA methylation (DNAm), is considered as ageing biomarker. We investigated the relationship between acceleration of six markers of the epigenetic age acceleration (EAA) and incident colorectal cancer during a 16-year follow-up in a nested case-control study from a population cohort of middle and older age in Russia.

Methods and Algorithms: A random population sample (Novosibirsk) was examined at baseline in 2003/05 (men, women, 45–69, n=9360) and followed up for 16 years. In a nested case-control design, we randomly selected 35 cases of CRC from an entire set of incident colorectal cancer (154), and 354 matched controls. The assessment of DNAm was conducted on Illumina EPIC850k Beadchip. Six baseline EA measures (Horvath, Hannum, PhenoAge, Skin and Blood (SB), BLUP, Elastic Net (EN)) were calculated, along with their respective EAA [2].

Results: DNAm ages calculated by Horvath's, Skin and Blood, BLUP and EN clocks were close to participants' chronological age (CA); the corresponding mean (SD) difference were -4.66 (5.25), -0.07 (4.08), 4.19 (3.16) and 0.30 (3.29). Hannum and PhenoAge EAs were less close to CA. The mean EAAs for four markers were significantly higher in CRC cases compared to control 3.91 (5.45) vs. -0.43(4.53), $p<0.001$; 2.15 (3.41) vs. -0.24(4.10) $p<0.001$; 2.29 (4.73) vs. -0.22(5.69), p-0.005; 1.83 (2.41) vs. -0.11(2.90) $p<0.001$ for Horvath, Hannum, PhenoAge, and BLUP, correspondingly. In multivariable adjusted logistic regression, the odds ratios (ORs) for CRC risk per decile increase in EAA ranged from 1.20 (95 % CI: 1.04-1.39) to 1.44 (95 % CI: 1.21-1.76) for Horvath, Hannum, PhenoAge, and BLUP measures. Conversely, the SB and EN EAA measures showed borderline inverse associations with CRC and had ORs of 0.86-0.87 (95 % CI: 0.76-0.99). Tertile analysis reinforced a positive association between CRC risk and four EAA measures (Horvath, Hannum, PhenoAge, BLUP) and a modest inverse relationship with EN EAA

Conclusion: In this case-control study based on a population cohort of middle and elderly age, we found a direct association between incident CRC and four markers of accelerated baseline epigenetic age (Horvath, Hannum, PhenoAge, and BLUP). In contrast, two markers showed a negative or no association. These results warrant further exploration in larger cohorts [3] and may have implications for CRC risk assessment and prevention.

Funding: The study is supported by Russian Science Foundation (No. 20-15-00371-П).

References

1. Dekker E. et al. Colorectal cancer. *Lancet.* 2019;394:1467-1480. doi 10.1016/S0140-6736(19)32319-0
2. Pelegí-Sisó D., De Prado P., Ronkainen J., Bustamante M., González J.R. Methylclock: a Bioconductor package to estimate DNA methylation age. *Bioinformatics.* 2021;37(12):1759-1760. doi 10.1093/bioinformatics/btaa825
3. Widayati T.A. et al. Open access-enabled evaluation of epigenetic age acceleration in colorectal cancer and development of a classifier with diagnostic potential. *Front Genet.* 2023;14:1258648. doi 10.3389/fgene.2023.1258648

doi 10.18699/bgrs2024-13-33

Transcriptional signatures and age-related changes in CD8+HLA-DR+ regulatory T cells

Matveeva K.[1]*, Kolmykov S.[2], Shevyrev D.[3]

[1] *Department of Immunobiology and Biomedicine, Sirius University of Science and Technology, Sirius, Russia*

[2] *Department of Computational Biology, Sirius University of Science and Technology, Sirius, Russia*

[3] *Resource Center of Cell Technology and Immunology, Sirius University of Science and Technology, Sirius, Russia*

** ksum.miha@gmail.com*

Key words: immunology; regulatory T cells; aging; single cell transcriptomics

Motivation and Aim: The subset of $CD8^+HLA\text{-}DR^+$ regulatory T lymphocytes ($CD8^+$Treg) was first described in 2014 [1]. These cells are known to suppress effector T cells with the involvement of checkpoint inhibitor molecules and have some properties of conventional $CD4^+$Treg [2]. Nevertheless, the nature and function of this subset remain poorly understood. Moreover, the study of the properties of this population becomes important in the context of general age-associated changes in the human immune system and higher vulnerability of $CD8^+$ T lymphocytes to these changes. In preliminary experiments, the $CD8^+HLA\text{-}DR^+$ population was found to segregate into two subpopulations based on surface expression of CD127 (IL7R). Based on this, the $CD8^+$Treg phenotype was defined as $CD3^+CD8^+HLA\text{-}DR^+CD127^{low}$. In this work, we aimed to determine transcriptional signatures and age-related changes in gene expression within the $CD8^+$Treg population using publicly available single-cell RNA-seq (scRNA-seq) data. Our objectives were as follows: (1) Identify the $CD8^+$Treg cluster; (2) Determine transcriptional signatures of $CD8^+$Treg; (3) Evaluate age-related gene expression changes in the $CD8^+$Treg cell population.

Methods and Algorithms: In this study, we analyzed peripheral blood mononuclear cells (PBMCs) scRNA-seq data from 982 donors, aged 19 to 97 y.o., from the OneK1K cohort [3] and 99 healthy donors, aged 22 to 75 y.o., from the SLE study [4]. The Scanpy package (version 1.9.8) [5] was used to analyze scRNA-seq data. Initially, data preprocessing steps such as normalization, logarithmization, scaling, dimensionality reduction (by PCA), and batch correction were performed. Subsequently, cells were clustered using the Leiden algorithm. The Celltypist package [6] was utilized to annotate the cell clusters and identify cluster of $CD8^+$ T effector cells ($CD8^+$Teff). Then, to more accurately represent CD8+Teff cells in reduced dimensionality embedding, the cells in a given cluster were returned logarithmized gene counts and the preprocessing steps from scaling to batch correction were performed repeatedly. Next, re-clustering and further manual annotation were conducted to identify the $CD8^+HLA\text{-}DR^+CD127^{low}$ cell subpopulation. The cell populations were pooled to conduct a pseudobulk differential expression analysis using edgeR (version 4.0.16) [7]. Functional analysis of the differentially expressed genes was performed using the g:Profiler web tool [8].

The described analyses are available as Jupyter-notebooks on a GitHub page: https://github.com/estakathersun/cd8treg_sc.

Results: To identify the signatures of the CD8$^+$Treg, data from donors under 60 years of age were utilized, as T cells at this age typically don't exhibit significant age-related changes. It's worth mentioning, these cells weren't segregated into a distinct cluster but distributed among the CD8$^+$Teff population. Therefore, CD8$^+$Teff cells displaying increased HLA-DR (HLA-DRA, HLA-DRB1 and HLA-DRB5 genes) and decreased IL7R expression levels were considered as the CD8$^+$Treg subpopulation. Differential expression analysis of the single-cell data revealed that the cell population, compared to the overall CD8$^+$Teff cell population, exhibited increased expression levels of genes associated with MHCII-mediated antigen presentation, cytotoxicity, and cytoskeleton organization (Fig. 1). However, at the pseudo-bulk level, no statistically significant differences were observed between the CD8$^+$Treg and CD8$^+$Teff populations.

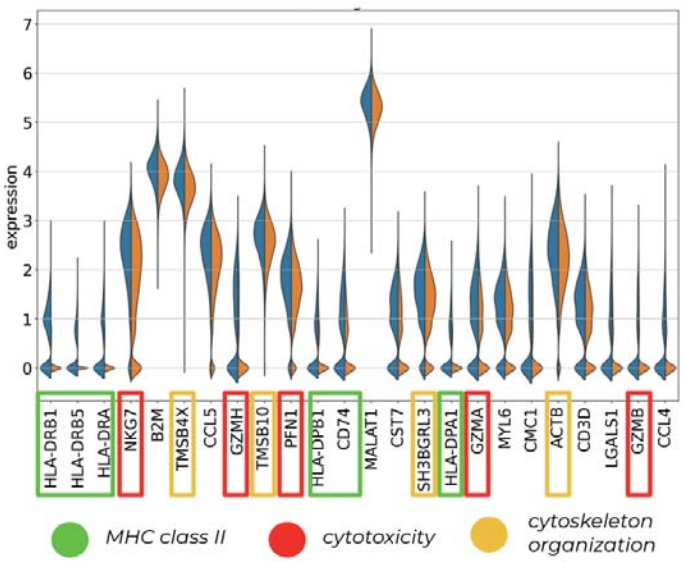


Fig. 1. Differential gene expression of CD8$^+$Treg (blue) in relation to the common CD8$^+$Teff population (orange). Colored boxes signify genes related to the corresponding function

Subsequently, we analyzed CD8$^+$Treg populations from two age groups: young adults, aged 20 to 35 years, and old individuals, aged 70 to 97 years. Fig. 2 shows the results of differential expression analysis at the pseudobulk level. A shift in the transcriptional profile of the studied cell population towards a terminally differentiated cell phenotype was observed. Notably, there was a decrease in the expression levels of certain cytotoxic protein genes, such as LYZ, GZMA and GZMK, while the expression of GZMH, typical of terminally differentiated CD8$^+$ T cells, increased. Additionally, we observed a positive association between age and increase in MHC-II expression level. However, there was no rise in the expression levels of genes associated with CD8$^+$ T cell exhaustion. It can be assumed that this cell population likely transitions towards a terminally differentiated phenotype with age but does not undergo cell exhaustion.

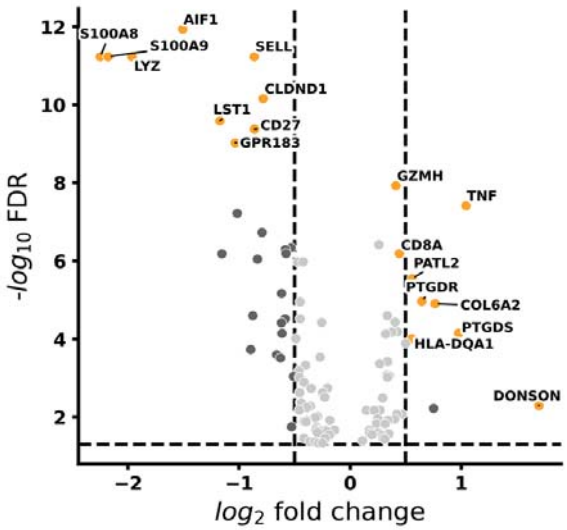


Fig. 2. Differentially expressed genes in the CD8$^+$Treg population from the old age group (70–97 y.o.)

Conclusion: CD8$^+$Treg is a heterogeneous subpopulation of effector CD8$^+$ T lymphocytes. CD8$^+$Treg tends to increase expression of genes associated with cytotoxicity, cytoskeletal rearrangement, and MHC II-mediated antigen presentation. This suggests that mechanism of CD8$^+$Treg-mediated suppression is cell-contact dependent cytolysis of target cells. In the old age group, the CD8$^+$Treg shows a change in marker gene expression to a terminal differentiated cell phenotype.

Funding: The study was financially supported by the Russian Science Foundation (project № 24-15-20003, https://rscf.ru/en/project/24-15-20003/).

References

1. Arruvito L., Payaslián F., Baz P. et al. Identification and Clinical Relevance of Naturally Occurring Human CD8+HLA-DR+ Regulatory T Cells. *J Immunol.* 2014;193(9):4469-4476
2. Machicote A., Belén S., Baz P., Billordo L.A., Fainboim L. Human CD8+HLA-DR+ Regulatory T Cells, Similarly to Classical CD4+Foxp3+ Cells, Suppress Immune Responses via PD-1/PD-L1 Axis. *Front Immunol*. 2018;9:2788
3. Yazar S., Alquicira-Hernandez J., Wing K. et al. Single-cell eQTL mapping identifies cell type – specific genetic control of autoimmune disease. *Science*. 2022;376(6589):eabf3041
4. Perez R.K., Gordon M.G., Subramaniam M. et al. Single-cell RNA-seq reveals cell type – specific molecular and genetic associations to lupus. *Science*. 2022;376(6589):eabf1970
5. Wolf F.A., Angerer P., Theis F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19(1):15
6. Domínguez Conde C., Xu C., Jarvis L.B. et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*. 2022;376(6594):eabl5197
7. Chen Y., Chen L., Lun A.T.L., Baldoni P.L., Smyth G.K. edgeR 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *bioRxiv*. 2024
8. Kolberg L., Raudvere U., Kuzmin I., Adler P., Vilo J., Peterson H. g:Profiler – interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* 2023;51(W1):W207-W212

doi 10.18699/bgrs2024-13-34

Common mechanisms of Alzheimer's disease and age-related macular degeneration: focus on MAPK signaling pathways

Muraleva N.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** myraleva@bionet.nsc.ru*

Key words: aging; neurodegeneration; Alzheimer's disease; age-related macular degeneration; MAPK signaling pathway; phosphorylation; OXYS rats

Motivation and Aim: Age is the main risk factor for Alzheimer's disease (AD) as the most common progressive senile dementia, and age-related macular degeneration (AMD), which is the leading cause of vision loss in people over 60 years of age. There are no effective ways to prevent and treat these neurodegenerative diseases, due to the incomplete knowledge of their pathogenesis. Risk factors (smoking, hypertension, hypercholesterolemia, atherosclerosis, obesity) and the mechanisms of the pathogenesis of AD and AMD are intersecting. Among them, oxidative stress, inflammation, mitochondrial dysfunctions and impaired maintenance of proteostasis with the pathological aggregation and accumulation of abnormal extracellular deposits – senile plaques in the brain of patients with AD, and in the drusen of patients with AMD. The development of these abnormalities may be associated with alteration of MAPK signaling pathways regulation with age. The contribution of impaired MAPK signalings to the pathogenesis of AD has been confirmed by the results of numerous investigations. MAPK is considered as a potential target for therapeutic interventions. Nevertheless, the information about their changes with age is limited. There are practically no data on the state of MAPK signaling activity in the retina during the development of AMD. The aim of this study was to compare changes in the activity of ERK1/2, p38MAPK and JNK signaling pathways in brain structures and retina with age and during the development of AD and AMD. We used Wistar (control) and OXYS rats as the models of accelerated aging, one of the manifestations of which is the simultaneous development of signs of AD and AMD.
Methods and Algorithms: The RNA-seq data obtained for prefrontal cortex, hippocampus and retina of Wistar and OXYS rats at the age of 20 days (signs of AMD and AD-like pathology absent), 3–5 months (period of their manifestation) and 18 months (active progression) were used to analyze differentially expressed genes involved in ERK1/2, p38MAPK and JNK signaling pathways according to Rat Genome Database.
Results: It has been established that changes in gene expression, which, according to the Rat Genome Database, are involved in p38MAPK, EPK1/2 and JNK signaling pathways, are tissue-specific and depend on the animal genotypes. In the retina, the number of differentially expressed genes (DEGs) involved in p38MAPK, ERK1/2 and JNK decreased in Wistar and OXYS rats with age. The maximum number of DEGs were detected at the age of 20 days, at the “preclinical” stage of the development of AMD-like pathology in OXYS rats. Among them were both activators and inhibitors of the activity of MAPK signaling pathways. By age three months and beyond found signs of a decrease in the activity of signaling pathways at the level of gene expression. At these

ages, only one DEG involved in the JNK signaling pathway was identified in the OXYS rat retina.

In the brain, the pattern of age-related changes in the expression profile of genes involved in ERK1/2 and p38MAPK signaling pathways in Wistar and OXYS rats was similar to that observed in the retina. In the prefrontal cortex and hippocampus, the number of DEG increased with age, the expression of which differed in OXYS compared to Wistar rats.

The level of phosphorylation of the key kinases p38MAPK, ERK1/2 and JKN signaling pathways is an objective indicator of their activity. Next, we evaluated it at the protein level by Western blot analysis. In all studied tissues – in the retina, prefrontal cortex, and hippocampus, a significant increase in the content of phosphorylated forms of p38MAPK, ERK1/2 and JNK with age was found in rats of both strains. At the same time, in OXYS rats, their accumulation was more active and increased as the signs of AD and AMD progressed.

Thus, our results showed that during physiological aging of Wistar rats and during accelerated aging of OXYS rats in the retina and brain, the direction of changes in the activity of MAPK signaling cascades both at the level of gene expression and at the protein level is the same. At the same time, it was found that the manifestation and progression of signs of AD and AMD pathology in OXYS rats occur against the background of enhanced phosphorylation of p38MAPK, ERK1/2 and JNK which makes it possible to consider an increase in the activity of MAPK signaling pathways as a common mechanism for the development of AD and AMD already at the early stages of these diseases.

Conclusion: Thus, our results showed that during physiological aging of Wistar rats and during accelerated aging of OXYS rats in the retina and brain, the direction of changes in the activity of MAPK signaling cascades both at the level of gene expression and at the protein level is the same. At the same time, it was found that the manifestation and progression of signs of AD and AMD pathology in OXYS rats occur against the background of enhanced phosphorylation of p38MAPK, ERK1/2 and JNK which makes it possible to consider an increase in the activity of MAPK signaling pathways as a common mechanism for the development of AD and AMD already at the early stages of these diseases.

Funding: The study is supported by Russian Scientific Foundation Grant (No. 24-25-00030).

doi 10.18699/bgrs2024-13-35

Effect of age and long-term disuse on proteome of human skeletal muscle

Orlova M.A.[1]*, Zgoda V.G.[2], Vavilov N.E.[2], Lednev E.M.[1], Zhedyaev R.Yu.[1], Vepkhvadze T.F.[1], Efimenko A.Yu.[3], Popov D.V.[1]

[1] *Institute of Biomedical Problems, RAS, Moscow, Russia*

[2] *Institute of Biomedical Chemistry, Moscow, Russia*

[3] *Lomonosov Moscow State University, Moscow, Russia*

* *meera.orlova@gmail.com*

Key words: skeletal muscle; proteome; hypokinesia; ageing; chronic inflammation

Motivation and Aim: In normal conditions, skeletal muscle accounts for 30-40 % of body mass. The main function of skeletal muscles is to maintain posture and movement; in addition, with the normal level of physical activity, skeletal muscle plays a significant role in the regulation of carbohydrate and fat metabolism in the organism and positively affects the functioning of various systems (cardiovascular, nervous, etc.). Skeletal muscle tissue undergoes pronounced changes with aging, in particular, the amount of mitochondria and muscle mass decreases, which leads to a decline in strength, endurance, fat and carbohydrate oxidation rate, and insulin sensitivity. Chronically reduced physical activity is one of the key factors inducing these changes. The aim of this work is to investigate the effect of chronic decrease in physical activity (disuse) on the human skeletal muscle proteomic profile and to compare these changes with those induced by age.

Methods and Algorithms: In order to investigate the effects of ageing (the combined effects of 'normal' ageing and pathological processes), older adults (72[69-77] years, n=37) with primary knee/hip osteoarthritis, a model of aging with chronic disuse and inflammation, were compared with young healthy adults (35[28-38] years, n=15). The effects of chronic long-term disuse were studied by comparing young adults with primary knee/hip osteoarthritis (39[37-42] years; n=8) with young healthy individuals. Biopsy samples from *m. vastus lateralis* were taken from all volunteers for proteomic profile studies for panoramic quantitative (isobaric labelled) mass spectrometry based analysis.

Results: A total of 1900 proteins that were detected in all samples were predominantly cytoskeletal, sarcomeric and mitochondrial proteins, as well as regulators of carbohydrate and fat metabolism. In elderly patients, relative to young healthy controls, there was a decrease in muscle mass (40 %, $p<0.001$) and changes in the expression of more than 350 proteins. The most striking changes were a decrease in the relative content of mitochondrial respiratory enzymes and glycolysis enzymes, and an increase in the expression of proteins associated with nucleic acids, cytoskeleton and membrane. This is in good agreement with previous studies investigating age-related changes in the proteomic profile of human skeletal muscle [1-3]. Chronic (long-term) disuse led to a decrease in muscle mass (by 15 %, $p<0.001$) and altered the relative content of 178 proteins (predominantly increased expression of inflammatory response regulators). Notably, there was no detectable drop in the expression of mitochondrial proteins, which were most significantly decreased in elderly patients, relative to young healthy controls.

A decrease in mitochondrial proteins was found when comparing elderly and young patients – people of different ages with comparable levels of disuse and increased inflammatory markers.

Conclusion: In elderly patients, we observe large-scale changes in the proteome, associated mainly with a decrease in mitochondrial protein expression and an increase in inflammatory protein expression. These changes are induced by different factors: long-term disuse leads mainly to an increase in the expression of inflammatory proteins, whereas the decrease in mitochondrial proteins is mainly related to age.

Funding: the research was supported by the RNF grant No. 21-15-00405 the state assignment of Lomonosov Moscow State University (biomaterials).

References

1. Murgia M., Toniolo L., Nagaraj N. et al. Single Muscle Fiber Proteomics Reveals Fiber-Type-Specific Features of Human Muscle Aging. *Cell Rep*. 2017;19(11):2396-2409
2. Theron L., Gueugneau M., Coudy C. et al. Label-free quantitative protein profiling of vastus lateralis muscle during human aging. *Mol Cell Proteomics*. 2014;13(1):283-294
3. Ubaida-Mohien C., Lyashkov A., Gonzalez-Freire M. et al. Discovery proteomics in aging human skeletal muscle finds change in spliceosome, immunity, proteostasis and mitochondria. *Elife*. 2019;8:e49874

doi 10.18699/bgrs2024-13-36

Investigation of the retinal and cortical vascular barriers' permeability along the onset and progression of neurodegenerative pathologies in a rat model

Peunov D.A.*, Telegina D.V., Stefanova N.A.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** d.peunov@g.nsu.ru*

Key words: Alzheimer's disease; age-related macular degeneration; prefrontal cortex; retina; blood brain barrier; blood-retinal barrier OXYS rats; aging

Motivation and Aim: Age-related macular degeneration (AMD) and Alzheimer's disease (AD) are neurodegenerative disorders, serving the main cause for central blindness and dementia in modern aging population respectively. High comorbidity of the diseases also worsen condition of patients [1]. Both pathologies are known to be associated with a vasculature dysfunction and recently with abnormalities of blood-retinal (BRB) and blood brain (BBB) barriers. It's considered that healthy aging is accompanied by decline of BRB and BBB and they are disrupted in both AMD and AD. However it remains unclear whether barriers' dysfunction cause or follows development of the diseases. The purpose of this study was to trace changes in BRB and BBB permeability along onset and progression of AMD-like and AD-like pathology, possibly revealing some common patterns. *Methods and Algorithms*: The experiments was performed on senescence-accelerated OXYS rats, which simulate key characteristics of both sporadic AD and dry AMD [2]. During the periods, when clinical signs of AD and AMD in OXYS rats are absent (age 20 and 45 days), manifest (~5 months), and progress (at 12 and 18 months), we assessed the permeability of BRB BBB for Evans blue dye (EBD) and the retinal and prefrontal cortex (PFC) level of the tight-junctional proteins occludin (Ocln) and claudin-5 (Cldn5) by Western blot analysis as described early [3]. Age matched Wistar rats served as controls.
Results: Two-way ANOVA shown that in retina concentration of EBD depended on strain as well as age ($F_{1, 43}$=15.5, p<0,001; $F_{3, 43}$=4,8, p<0,01) and was lower in OXYS rats compared to the Wistar. Post-hoc test confirmed that at the age of 20 days, 5 and 18 month it was significantly lower (p<0.05). At the age of 18 month, observed only in OXYS rats decrease of EBD concentration (p<0.05) from the age of 12 month, was the cause (Fig. 1, A).
Since the data on the concentration of the EBD in PFC didn't met the assumption of normality we performed Kurskal-Wallis test followed with pair by pair t-test comparison. At the first stage we shown that the concentration of the dye in PFC was significantly dependent on age (Chi-squred$_3$=19.2, p<0.001) but not on strain. Post-hoc comparison revealed tendency in the OXYS rats at the age 1.5 month to have higher level of the EBD, compared to the Wistar. Both strains shown significant gradual decrease (p<0.05) of the EBD concentration from the age of 45 days until 18 month (p<0.05). In OXYS rats the tendency of the dye content to drop was detectable earlier than in Wistar (Fig. 1, B).

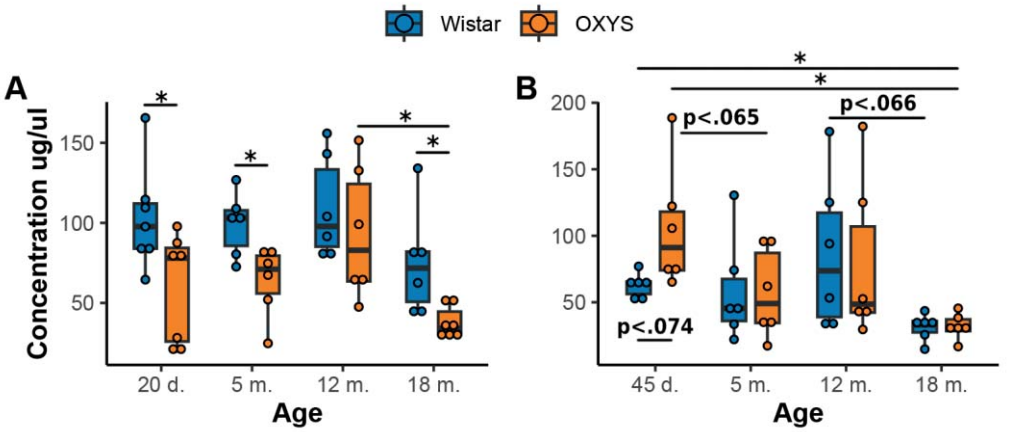


Fig. 1. Concentration of retained EBD in retina (A) and prefrontal cortex (B) of OXYS and Wistar rats along the progression of pathologies. * – significant at p<0.05. Boxes state for Q1–Q3 range, line for median, whiskers end on the min/max values. d – days, m – months

The regulation and restriction of the blood-to-tissue and reverse transports considered to be the key function of BBB and BRB and is carried out by tight junctions (TJ). Among others claudin-5 (Cldn5) and occludin (Ocln) are considered to be those of importance for TJ functioning in BRB and BBB. While Cldn5 is directly forms intercellular bonds, Ocln serves regulatory functions. The level of Cldn5 in retina depended on the strain ($F_{1,24}$=4.4, p<0.05), specifically tended to be higher in the OXYS compared to the Wistar rats. (fig. 2, A). The amount of Ocln in retina was higher in the OXYS rats relatively to the Wistar at the age of 18 month (p<0.05). It was caused by decrease from the age of 20 days, detected only in the Wistar rats (fig. 2, B). In PFC We haven't detected any changes nor in Cldn5 (fig. 3, A) neither Ocln levels (Fig. 3, B).

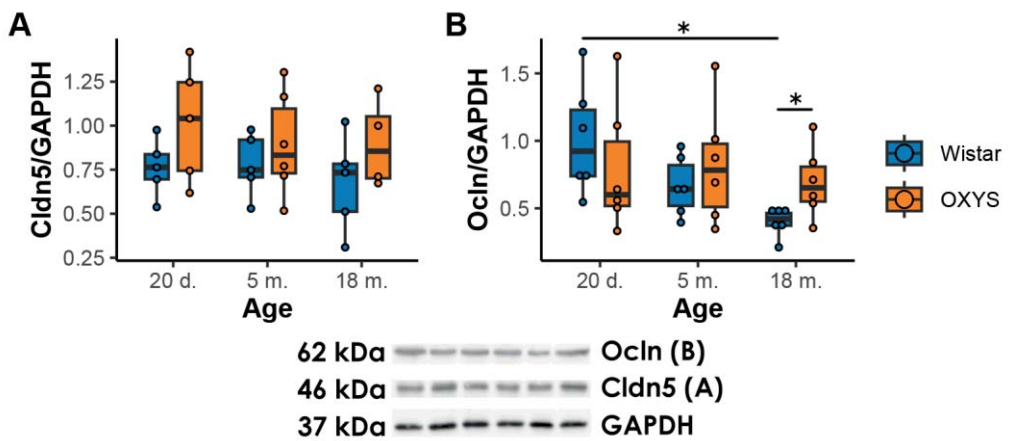


Fig. 2. level of Cldn5 (A) and Ocln (B) in retina of Wistar and OXYS rats along the progression of pathologies. * – significant at p<0.05. Boxes stand for Q1–Q3 range, line for median, whiskers end on the min/max values. d – days, m – months

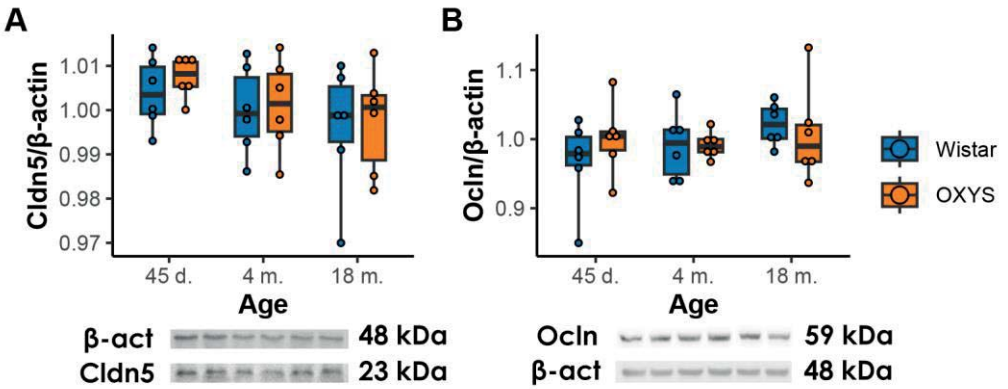


Fig. 3. level of Cldn5 (A) and Ocln (B)in cortex of Wistar and OXYS rats along the progression of pathologies. Boxes stand for Q1–Q3 range, line for median, whiskers end on the min/max values. d – days, m – months

Conclusion: According to our data, BRB and BBB permeability for the EBD weakly altered with age in both retina and PFC in Wistar and OXYS rats. But in PFC of OXYS rats the permeability for the dye tended to decrease earlier than in Wistar due to the higher level of BBB permeability at the age of 45 days in OXYS rats. In OXYS rats AMD-like retinopathy develops against the lowered permeability of the BRB, which may result from the decreased level of vascular growth and increased activity of Müller cells, which was shown earlier [4, 5]. Vascular growth factor deficit could dampen vesicular transport, which is an alternative way, how albumin enters the tissue. Thus it may impact transport of the albumin-binded EBD. Slight elevation in overall Cldn5 content in retina also could be explained with upregulated activity of Müller cells, as they are able to bring up expression of Cldn5 in retinal pigment epithelium, where it is normally absent, as a compensatory reaction to a pathology progression.

Overall the obtained data supports a role of BRB dysfunction in AMD progression and assumes that leakiness of barriers is not immanent to the AMD and AD development at least the portion could be measured with EBD.

Funding: The study is supported by State Budget Project (No. FWNR-2022-0016).

References

1. Wen L.-Y. et al. Increased risk of Alzheimer's disease among patients with age-related macular degeneration: A nationwide population-based study. *PLoS One*. 2021;16(5):e0250440
2. Stefanova N. et al. Senescence-accelerated OXYS rats: a model of age-related cognitive decline with relevance to abnormalities in Alzheimer disease. *Cell Cycle*. 2014;13(6):898-909
3. Telegina D.V. et al. Evaluation of the State of the Blood-Retinal Barrier during the Development of Signs of Age-Related Macular Degeneration in OXYS Rats. *Moscow University Biological Sciences Bulletin*. 2023;78(3):190-197
4. Kozhevnikova O.S. et al. VEGF and PEDF levels in the rat retina: effects of aging and AMD-like retinopathy. *Adv Gerontol*. 2018;31(3):339-334
5. Telegina D.V. et al. Contributions of age-related alterations of the retinal pigment epithelium and of glia to the AMD-like pathology in OXYS rats. *Sci Rep*. 2017;7(1):41533

doi 10.18699/bgrs2024-13-37

Insufficient neuronal support by glial cells as a predictor and driver of Alzheimer's disease-like pathology in OXYS rats

Stefanova N.A.[1]*, Rudnitskaya E.A.[1, 2], Burnyasheva A.O.[1], Kozlova T.A.[1], Kolosova N.G.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Ben-Gurion University of the Negev, Beer Sheva, Israel*

* *stefanovan@mail.ru*

Key words: Alzheimer's disease; astrocyte; microglia; OXYS rats

Motivation and Aim: Alzheimer's disease (AD) is a multifactorial neurodegenerative disorder and the most common cause of dementia in an aging population. Prevalence of AD dramatically increases in people aged 65 years or older. To date, there is no known medication that will stop the course of AD, which is largely due to incomplete knowledge of disease pathogenesis. Recently more attention has been given to the role of glial cells, especially microglia and astrocytes, in the process of neurodegeneration. Astrocytes are the most common type of cells in the central nervous system that perform many functions in the brain including neuronal protection and maintenance of homeostasis. Microglia are the main resident immune cells of the brain, which are involved in homeostasis and in host defense against pathogens and disorders of the central nervous system. In various pathological conditions, including AD, they go into states associated with either an increase or loss of their functions, which contributes to neuroinflammation and neurodegeneration. Currently, the influence and support of neurons by glial cells is considered as one of the possible strategies for slowing the development of the disease. Today, it is obvious that inappropriate support of neurons by glial cells plays an important role in the pathophysiology of AD. However, it is not clear whether it plays a key role in the development of the disease or is a consequence of its progression?

Methods and Algorithms: Using senescence-accelerated OXYS rats as a suitable model of AD [1] we examined the features of glial support in the hippocampus and frontal cortex during the whole life of animals: from the birth to the progressive stages of AD-like pathology. Wistar rats were used as healthy controls.

Results: We found that OXYS rats are born with lower density of astrocytes and microglial cells and higher extent of apoptosis in the hippocampus and frontal cortex compared to control Wistar rats [2]. This situation persists for all first postnatal week with the one exception at postnatal day 3: at this time point, astrocyte density in the hippocampus increased and became higher in OXYS rats than in Wistar rats, which together with significant increase of apoptosis may accompany increased neurogenesis in hippocampal dentate gyrus (DG) [2]. As for the functional state of glial cells, it was worse in OXYS rats in the first week after birth. At the second week of life the glial support normalized, and at the end of the period of brain maturation–that means, the end of the third week of life–astrocyte density became even higher in the hippocampus of OXYS rats with simultaneous increase in the apoptosis extent and intensity of neurogenesis in DG [3]. To the end of adolescence–that means, to the age of 1.5 months–astrocytic density normalized, whereas the density of microglia and, especially, its

activated fraction was higher in the hippocampus of OXYS rats compared to Wistar rats. The first signs of neurodegeneration in OXYS rats are observed at the age of 3–5 months. Interestingly, this period is not accompanied by any significant changes in glial support of the hippocampus or frontal cortex of OXYS rats. The period of amyloid-β accumulation – that means, 12 months of age – is accompanied by increased neuro- and gliogenesis in the hippocampus of OXYS rats as well as by the tendency to higher microglia density. This activation may be of compensatory nature and direct to slow down neurodegenerative processes. Activated gliogenesis at 12 months of age led to increased astrocyte density in the hippocampus at 18 months of age – at the progressive stage of AD-like pathology in OXYS rats. This increase in astrocyte density during the progression of AD signs in OXYS rats can be considered as support to the neurons and synapses against degeneration or as the sign of astrogliosis accompanying neurodegenerative changes in the hippocampus.

Different interventions can affect the course of neurodegeneration through their influence on glial support of the neurons and synapses. We examined the effectiveness of cognitive training and systemic antioxidant SkQ1 administration to normalize glial support of hippocampus of OXYS rats at the progressive stage of AD-like pathology (18 months of age). While cognitive training even increased the already higher density of astrocytes in OXYS rats, SkQ1 administration did not affect this parameter and significantly decreased the density of activated microglia cells thus exerting anti-inflammatory effect [4].

Conclusion: In general, our results indicate that the most prominent alterations in glial support of hippocampus and prefrontal cortex of OXYS rats occur in early life, at the course of brain development and maturation. These alterations contribute to neurodegenerative processes and the development of AD signs and may serve as early predictors of neurodegeneration. At the same time, altered glial support at the progressive stages of AD-like pathology in OXYS rats is more likely reactive rather than causative. Different interventions may improve glial support and as a consequence slow down neurodegeneration, however the proper time for these interventions need to be determined.

Funding: The study is supported by basic project No. FWNR-2022-0016.

References

1. Stefanova N.A., Kozhevnikova O.S., Vitovtov A.O. et al. Senescence-accelerated OXYS rats: A model of age-related cognitive decline with relevance to abnormalities in Alzheimer disease. *Cell Cycle.* 2014;13(6):898-909. doi 10.4161/cc.28255
2. Rudnitskaya E.A., Kozlova T.A., Burnyasheva A.O., Stefanova N.A., Kolosova N.G. Glia Not Neurons: Uncovering Brain Dysmaturation in a Rat Model of Alzheimer's Disease. *Biomedicines.* 2021;9(7):823. doi 10.3390/biomedicines9070823
3. Rudnitskaya E.A., Kozlova T.A., Burnyasheva A.O. et al. Features of Postnatal Hippocampal Development in a Rat Model of Sporadic Alzheimer's Disease. *Front Neurosci.* 2020;14:533. doi 10.3389/fnins.2020.00533
4. Rudnitskaya E.A., Burnyasheva A.O., Kozlova T.A., Peunov D.A., Kolosova N.G., Stefanova N.A. Changes in Glial Support of the Hippocampus during the Development of an Alzheimer's Disease-like Pathology and Their Correction by Mitochondria-Targeted Antioxidant SkQ1. *Int J Mol Sci.* 2022;23(3):1134. doi 10.3390/ijms23031134

doi 10.18699/bgrs2024-13-38

Changes in behavior and the stress-responsive system in the aging SHR and WKY rats after prolonged individual housing

Stepanichev M.*, Mamedova D., Nedogreeva O., Ovchinnikova V., Moiseeva Yu., Tret'yakova L., Gulyaeva N.

Institute of Higher Nervous Activity and Neurophysiology, RAS, Moscow, Russia

** m_stepanichev@ihna.ru*

Key words: aging; social isolation; hypertension; spontaneously hypertensive rats; Wistar-Kyoto rats; anxiety; depression; social preference and novelty; corticosterone

Motivation and Aim: Social isolation or loneliness in elderly is a serious problem in many developed and developing countries in the world including Russian Federation. Social isolation has objective and subjective dimensions that significantly influences its consequences for each individual. Prolonged social isolation is believed to negatively impact mental health, inducing psychiatric impairments and disturbances of cognitive functions. In older adults, the effects of isolation on mental health may significantly depend on age of a person and be strongly associated with individual perception of isolation. Furthermore, aging is associated with multimorbidity and cardiovascular and cerebrovascular diseases are among the most often reasons of morbidity and mortality in this group of people. The aim of the present study was to investigate the effects of prolonged social isolation on anxiety, depressive-like, and social behavior in the aging spontaneously hypertensive rats (SHR) and normotensive Wistar Kyoto (WKY). We also studied some biochemical indices of the stress-responsive system in these rats during and after the end of the isolation period.

Methods and Algorithms: Male 9–10-month-old SHR and WKY rats were used for the study. Animals were distributed into 4 groups according to their body weight and arterial pressure. The SHRi and WKYi rats were individually housed in opaque cages whereas the SHRs and WKYs rats remained in their home cages in groups of 2–3 animals. Individual housing continued for 12 weeks before the start of and then during the behavioral testing. Then, the animals were studied in the following battery of tests: sucrose preference test, open field test, elevated plus maze (EPM), three-chamber social test, and Barnes maze. The latter test was used to study learning ability and cognitive flexibility. The level of corticosterone was measured in the peripheral blood using ELISA. The expression of corticosterone-related genes was studied in the adrenals.

Results: Like in many other studies, increased locomotor and exploratory activity was revealed in SHR compared to WKY rats at the age of 13 months. The SHR rats performed increased number of open arm visits and time spent there as well as decreased stretching and higher risk-assessment behavior. However, we cannot suggest that the differences in behavior of animals in the EPM are determined by strain-specific trait/state anxiety only: rather they are associated with the hyperactivity of SHRs. This is supported by higher number of total entries into the maze arms. Surprisingly, social isolation induced only mild effect on anxiety in the open field test without influencing other affective-related behaviors. No effects of isolated housing on sociability or social

novelty preference were revealed. In contrast to emotional features, isolation induced some cognitive impairments observed during learning in the Barnes maze. WKYs WKYi, and SHRs rats were able to learn the spatial task in the maze whereas SHRi rats were bad learners. Moreover, WKYi, SHRs and SHRi animals could not relearn the location of the safety box after its moving into the opposite position. Despite the adrenal gland hypertrophy in the SHRs, corticosterone levels remained stable within the period of isolation but the expression of nuclear glucocorticoid receptor (*Nr3c1*) mRNA in the adrenals was lower in the SHR as compared to WKY rats.

Conclusion: We can assume that in the aging SHR rats long-term social isolation significantly impaired cognitive abilities. Social isolation negatively influenced cognitive flexibility in both normotensive and hypertensive animals. Thus, hypertension significantly contributed to the effects of social isolation on cognitive functions but not on depressive-like behavior. Further, these impairments are not directly related to functions of the stress-responsive system.

Funding: The study is supported by the RSF (No. 22-15-00132).

doi 10.18699/bgrs2024-13-39

Time-of-the-day–dependent alterations of glutamate/GABA system in hippocampus of rats with age and during of the Alzheimer's disease signs development

Telegina D.V.*, Kozhevnikova O.S., Kolosova N.G.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** telegina@bionet.nsc.ru*

Key words: glutamate; GABA; circadian; hippocampus; aging; Alzheimer's disease; OXYS rats

Motivation and Aim: Growing evidence suggests close correlations between Alzheimer's disease (AD) and circadian rhythm disruption. Along with cognitive impairment, circadian rhythm dysfunction is a main sing of AD. The hippocampus is a subordinate circadian oscillator. Here, more than 10 % of genes and proteins show circadian fluctuations and are associated with changes in synaptic excitability [1]. The study of circadian rhythms in the hippocampus during the development of AD is also beginning to gain momentum. Dysregulation of circadian clocks in the brain not only associated with clock genes, it may also affect neurotransmitter systems such as glutamate/GABA system. Glutamate and GABA are major excitatory and inhibitory neurotransmitters in the brain, and play a crucial role in the functioning of the hippocampus, a brain region important for learning and memory processes. But there are few studies on the diurnal variability of the glutamate/GABA system and most only analyze its variation in the suprachiasmatic nucleus [2]. There is little information on diurnal changes in the glutamatergic and GABAergic systems during hippocampus aging and even less at different stages of AD development.

In this research we analyzed the day-night variations the level of main proteins of glutamatergic and GABAergic systems in the whole hippocampus of the Wistar rats and of the senescence-accelerated OXYS rats, a model of the most common (>95 %) sporadic AD.

Methods and Algorithms: The work was carried out in the prematurely aging OXYS rats: a unique model of the sporadic form of AD. In these animals, all of the key signs of AD develop spontaneously in the absence of mutations in the *Psen1, App,* and *Psen2* genes typical of hereditary AD [3]. Animals were examined during several key periods: 20 days group corresponds to the "preclinical" period preceding the development of AD signs, during their manifestation (3 months), and active progression (18 months). Wistar rats of the same age were used as control. Animals were kept under the standard conditions of vivarium with light cycle 12-h light/12-h darkness; they received granulated feed and water *ad libitum*. Based on early data of diurnal rhythms of concentration of glutamate and GABA in rat's hippocampus, we explored two time periods – nighttime between 02.00–04.00 am and daytime between 10.00–12.00. In these time we investigeted the protein level of of glutaminase, glutaminesynthetase, GLAST, GLT1, NMDAr1, NMDA2B, GluA1, mGluR1, GAD67, GABA-T, GABAAr1, GAT1 by Western blot.

Results: Our data showed that some components of the glutamate- and GABAergic system are characterized by an increase at night and a decrease during the day in the

hippocampus of Wistar and OXYS rats at the all ages. So we detected more than 2-fold increase in subunits of NMDA receptor (subunits NR1 and NR2B; also known as GluN2B and GluN1, respectively), AMPA receptor (subunit 1 also known as GluA1) and GABA receptor (subunit alpha-1 also known as GABAAR1). Also it was observed increase at nighttime the protein level of the transporter that removes GABA (GAT1) and glutamate (GLAST and GLT1) from the synaptic cleft. While the levels of synthesis and degradation enzymes of GABA (GAD67 and GABA-T, respectively) and glutamate (glutaminase and glutamine synthetises, respectively) did not depend on the time of day in both Wistar and OXYS rats. It is important to note that protein level of NMDA and AMPA receptor subunits decreased at the nighttime in OXYS rats on the early and late stage of AD as compared Wistar rats.

Conclusion: The level of ionotropic receptor subunits and transporters of glutamate and GABA in hippocampus of Wistar and OXYS rats show variations depending on the time of day – increasing at the nigttime and decreasing at the daytime. Since these components of glutamate and GABA system are associated with synaptic plasticity, we believe that NMDA, AMPA, GABAA receptors and GABA and glutamate transporter is associated with circadian modulation of neurons and astrocytes and synaptic integration and hippocampal-dependent learning. Our data indicated that manifestation and progression of AD-like degeneration in OXYS rats accompanied by GluN2B, GluN1 and GluA1 level decreasing at the nighttime.

Funding: The study is supported by State Budget Project No. FWNR-2022-0016.

References

1. von Gall C. The effects of light and the circadian system on rhythmic brain function. *IJMS.* 2022;23(5):2778. doi 10.3390/ijms23052778
2. Estrada-Rojo F., Escobar C., Navarro L. Circadian variations of neurotransmitters in the brain – its importance for neuroprotection. *Revista Mexicana Neurociencia.* 2020;21(1):31-38. doi 10.24875/rmn.19000069
3. Kolosova N.G., Kozhevnikova O.S., Muraleva N.A., Rudnitskaya E.A., Rumyantseva Y.V., Stefanova N.A., Telegina D.V., Tyumentsev M.A., Fursova A.Z. SkQ1 as a tool for controlling accelerated senescence program: Experiments with OXYS rats. *Biochemistry (Moscow).* 2022;87(12):1552-1562. doi 10.1134/S0006297922120124

doi 10.18699/bgrs2024-13-40

Daily dynamics of autophagy gene expression in the rat retina during the development of retinopathy

Timofeeva Yu.V.[1, 2]*, Telegina D.V.[1], Devyatkin V.A.[1], Kolosova N.G.[1], Kozhevnikova O.S.[1]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *timofeeva.2002@inbox.ru*

Key words: circadian rhythms; retina; autophagy

Motivation and Aim: Age-related macular degeneration (AMD) is a complex neurodegenerative progressive eye disease, resulting in severe loss of central vision in the aging population. AMD pathogenesis is not completely known, but strongly associate with autophagy disruption. Autophagy is a conserved cellular degradation pathway for the breakdown of cytoplasmic components: damaged proteins and organelles. Besides a housekeeping function, autophagy is crucial for response to the stress. Defects in the autophagy are linked to aging and disease pathology. Circadian rhythms are a hallmark of physiology, but how such daily rhythms organize cellular catabolism is poorly understood. The retina has an own circadian system in that all retinal cells have their own oscillators which integrated for the rhythmic regulation of retinal physiology and function [1]. In disease states, dysregulation of the retinal circadian rhythms might further exacerbate the development of retinopathy. The daily autophagy demands in the RPE require precise gene regulation for the digestion and recycling of intracellular and POS components in lysosomes in response to light and stress conditions. The time course of daily fluctuations in the levels of autophagy occurring in the photoreceptors and RPE has not been characterized, and the mechanisms underlying the dynamic regulation of autophagy in these cells remains unknown. The purpose of this study was to determine the daily pattern of autophagy genes expression in the retina of young and old Wistar and senescence-accelerated OXYS rats which spontaneously reproduce the major signs of AMD: dystrophic alterations of the RPE, thinning of the neuroretina, and impairment of choroidal microcirculation [2].

Methods and Algorithms: Retinal tissues from 3-month and 21-month-old OXYS rats developing signs of AMD and Wistar rats ($n = 5$) were collected at ZT1, ZT8 and ZT16 in Zeitgeber time units, where ZT0 represents lights on (8:00 am) and ZT12 represents lights off (8:00 pm). Levels of autophagy markers expression (*Atg5, Atg7, Becn1, Gabarapl1, Nbr1, Map1lc3b, p62,* and *Ulk1*) were evaluated by qPCR using 3 reference genes (*Actb*, *Gapdh*, *Hprt1*).

Results: Changes in the mRNA levels of three key autophagy genes (*Atg5*, *p62*, *Ulk1*) were determined in relation to the time of day. For the *Atg7* gene, the level of *Atg7* in the retina depended on the age of the animals according to ANOVA ($p = 0.02$): post-hoc analysis showed that at ZT1, *Atg7* levels were higher in the retina of old OXYS rats than in young rats ($p = 0.0006$). In terms of daily dynamics, *Atg7* levels were lower at ZT8 and ZT16 than at ZT1 ($p = 0.002$ and $p = 0.009$ respectively) in old OXYS rats, but not in Wistar rats. Moreover, at ZT1, the retinal level of *Atg7* was higher in OXYS rats than in Wistar rats ($p = 0.04$). For the *p62* gene, ANOVA showed that *p62* mRNA levels

depended on the age of the rats ($p = 0.005$), with a comparison of group means showing significant differences between ages only for OXYS rats: at the time points ZT1 and ZT16, old rats had higher levels of *p62* than at 3 months of age ($p = 0.04$ and $p = 0.03$, respectively). We also observed significant diurnal changes in *p62* gene expression – a decrease at ZT8 compared to ZT1 ($p = 0.003$), and an increase at ZT16 compared to ZT8 ($p = 0.02$). In the case of the *Ulk1* gene, the level of *Ulk1* depended on the interaction of the factors genotype*age*time ($p = 0.04$). When comparing group means, significant differences were found between strains at 21 months of age at ZT8 ($p = 0.03$), with higher expression of the *Ulk1* gene in Wistar rats compared to OXYS rats. Post-hoc analysis showed significant diurnal changes in *Ulk1* gene expression: the level of *Ulk1* in old OXYS rats at was lower at ZT8 than at ZT1 ($p = 0.006$), at ZT16 the gene expression was higher than at ZT8 ($p = 0.048$).

Conclusion: At 21 months of age, OXYS rats have altered diurnal dynamics of expression of three key autophagy genes (*Atg7*, *p62*, *Ulk1*) in the retina compared to both Wistar rats of the same age and 3-month-old OXYS rats. No time-of-day (ZT1, ZT8, ZT16) or age (3-month-old rats and 21-month-old rats) related changes in autophagy gene expression (*Atg5*, *Atg7*, *Becn1*, *Gabarapl1*, *p62*, *Nbr1*, *Ulk1*) were detected in the retina of Wistar rats. These data may suggest that autophagy dysregulation occurs in the retina of prematurely ageing OXYS rats, which may lead to the progression of retinopathy development. Our study emphasizes the importance of the autophagic pathway in the pathogenesis of AMD and supports the view that disruption of circadian rhythms may influence retinal ageing.

Funding: The study is supported by State Budget Project No. FWNR-2022-0016.

References

1. Yao J. et al. Circadian and Noncircadian Modulation of Autophagy in Photoreceptors and Retinal Pigment Epithelium. *Invest Ophthalmol Vis Sci.* 2014;55(5):3237-3246. doi 10.1167/iovs.13-13336
2. Kolosova N.G. et al. SkQ1 as a Tool for Controlling Accelerated Senescence Program: Experiments with OXYS Rats. *Biochemistry (Moscow).* 2022;87(12):1552-1562. doi 10.1134/S0006297922120124

doi 10.18699/bgrs2024-13-41

Predicting the risk of all-cause mortality in patients with type 2 diabetes mellitus

Vershinina O.[1]*, Kondakova E.[1], Sabbatinelli J.[2, 3], Colombaretti D.[2], Bonfigli A.R.[4], Franceschi C.[1], Ivanchenko M.[1], Olivieri F.[2, 5]

[1] *Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia*

[2] *Department of Clinical and Molecular Sciences, Università Politecnica delle Marche, Ancona, Italy*

[3] *Clinic of Laboratory and Precision Medicine, IRCCS INRCA, Ancona, Italy*

[4] *Scientific Direction, IRCCS INRCA, Ancona, Italy*

[5] *Advanced Technology Center for Aging Research, IRCCS INRCA, Ancona, Italy*

* *olya.vershinina@itmm.unn.ru*

Key words: diabetes; mortality risk; machine learning; artificial intelligence; explainability

Motivation and Aim: Diabetes mellitus is a non-infectious endocrine disease characterized by high blood sugar levels. According to the International Diabetes Federation, its prevalence is steadily increasing: in 2021, there were approximately 536.6 million cases of diabetes diagnosed among people aged 20–79 years, and by 2045 this number is estimated to increase to 783.2 million people [1]. Diabetes leads to serious complications and increased mortality – diabetes was responsible for 6.7 million deaths worldwide in 2021. To improve the life quality and expectancy of patients with diabetes, it is necessary to develop personalized treatment strategies. Information about individual mortality risk predictions may be of great importance in choosing appropriate medications and lifestyle recommendations. Based on machine learning algorithms and explainable artificial intelligence, we develop an explainable model for predicting the mortality risk of patients with diabetes and identify important predictors that influence life expectancy.

Methods and Algorithms: We consider a dataset of 568 patients with type 2 diabetes mellitus (T2DM) who were followed at the Italian National Research Center on Aging (Ancona, Italy) for 16.8 years [2]. Among the patients, there were 308 men and 260 women aged from 40 to 87 years, the average age of patients was 65.74 ± 8.15 years. The dataset includes more than 100 different demographic, laboratory and clinical characteristics (features) that are considered risk factors for mortality. The data was pre-processed and divided into training and test sets. In constructing the predictive model, we considered the two most widely used survival models – Cox proportional hazards regression [3] and random survival forests [4]. In order to obtain an optimal and small subset of the most significant features, different feature selection strategies were used with each survival model: feature selection based on univariate Cox regression, forward selection method, backward selection method, selection based on feature importance. The quality of mortality risk prediction models was assessed using two metrics: Harrell's C-index and area under the receiver operating characteristic curve (AUC). Both quality metrics can take values from 0 (poor prediction) to 1 (perfect prediction). In order to increase user confidence in the final prediction model, an explainable artificial intelligence approach was used – SHapley Additive exPlanations (SHAP) [5], which allows interpreting model predictions both globally at the level of the entire sample and locally at the level of specific patients.

Results: As a result of experiments with different feature selection strategies and different survival algorithms, more than ten models were obtained that predict the mortality risks of patients with T2DM. It is interesting to note that all the resulting feature subsets included the age of the patients as one of the most important characteristics affecting life expectancy. More accurate models were obtained using random survival forests, for which the quality metric of the 15-year prediction on the test set varied within the AUC range from 0.77 to 0.82. The model with the highest AUC value was selected as the final model. The risk of mortality predicted by the model was significantly associated with survival, which was also confirmed by constructing Kaplan–Meier survival curves and using the log-rank test. Using the SHAP method, graphs of the global importance of features were constructed, as well as graphs of local explainability, which reflect which features increase the risk assessment of specific patients, and which, on the contrary, decrease.
Conclusion: Thus, an explainable machine learning model was developed to predict the mortality risks of patients with type 2 diabetes, which has high quality scores on both the training and test sets. The advantage of the model is the ability to assess the contribution of features to predictions and explain the predictions made.

References

1. Sun H. et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract*. 2022;183:109119. doi 10.1016/j.diabres.2021.109119
2. Bonfigli A.R. et al. Leukocyte telomere length and mortality risk in patients with type 2 diabetes. *Oncotarget*. 2016;7(32):50835-50844. doi 10.18632/oncotarget.1061
3. Cox D.R. Analysis of Survival Data. Chapman and Hall/CRC, 1984
4. Ishwaran H., Kogalur U.B., Blackstone E.H., Lauer M.S. Random survival forests. *Ann Appl Stat.* 2008;2(3):841-860. doi 10.1214/08-AOAS169
5. Lundberg S.M., Lee S.I. A unified approach to interpreting model predictions. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. 2017;4765-4774

doi 10.18699/bgrs2024-airus

Авторский указатель

doi 10.18699/bgrs2024-aieng

Author index

Научное издание

БИОИНФОРМАТИКА РЕГУЛЯЦИИ И СТРУКТУРЫ ГЕНОМОВ / СИСТЕМНАЯ БИОЛОГИЯ

Четырнадцатая международная мультиконференция

Тезисы докладов

Публикуется в авторской редакции

BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE / SYSTEMS BIOLOGY (BGRS/SB-2024)

The Fourteenth International Multiconference

Abstracts

Printed without editing

Выпуск подготовлен информационно-издательским отделом ИЦиГ СО РАН

Подписано к печати 02.08.2024. Сетевое издание

Федеральный исследовательский центр
«Институт цитологии и генетики Сибирского отделения Российской академии наук»
630090, Новосибирск, проспект Академика Лаврентьева, 10